08773-88-T

AFOSR - TR - 71 - 2612

AD 730925

HUMAN PERFORMANCE CENTER

DEPARTMENT OF PSYCHOLOGY

The University of Michigan, Ann Arbor

Theoretical Implications of Proactive Interference in Short-Term Memory

RAYMOND W. BENNETT

D D C
RECEIVED
OCT 15 1971
RECEIVED
C

Approved for public release;
distribution unlimited.

THE UNIVERSITY OF MICHIGAN
ARTES · SCIENTIA · VERITAS
1817

Technical Report No. 33

August 1971

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

100

BEST
AVAILABLE COPY

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) University of Michigan, Human Performance Center Department of Psychology, Ann Arbor, Michigan	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

THEORETICAL IMPLICATIONS OF PROACTIVE INTERFERENCE IN SHORT-TERM MEMORY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Raymond W. Bennett

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August, 1971	91 plus vii	27

8a. CONTRACT OR GRANT NO. AF 49(638)-1736 b. PROJECT NO. A.O. 461-11 c. 61101D d. 681313	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 33 08773-88-T 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR - TR - 71 - 2612

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES TECH,OTHER	12. SPONSORING MILITARY ACTIVITY AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (NL) 1400 WILSON BLVD ARLINGTON, VIRGINIA 22209

13. ABSTRACT

The primary goal of this study was to determine the extent to which confusions among traces are responsible for forgetting, most especially, for forgetting attributable to proactive interference. The experiment used a modification of the Brown-Peterson technique. The major alteration was to replace the recall test with a two-alternative forced choice recognition test. One of the recognition alternatives (the target) was one of the elements of the to-be-remembered-item (TBRI) and the other (the foil) was an element from some prior TBRI or it was a word not previously presented in the experiment. The major experimental manipulations were the recency of the foils and the length of the retention interval. Recognition accuracy was found to decrease as the recency of the foil increased and as the retention interval was lengthened. It is argued that this observation is sufficient to make implausible any model not assuming confusion among traces. A reasonably successful attempt was made to formulate a quantitative structure to handle these data.

THE HUMAN PERFORMANCE CENTER

DEPARTMENT OF PSYCHOLOGY

The Human Performance Center is a federation of research programs whose emphasis is on man as a processor of information. Topics under study include perception, attention, verbal learning and behavior, short- and long-term memory, choice and decision processes, and learning and performance in simple and complex skills. The integrating concept is the quantitative description, and theory, of man's performance capabilities and limitations and the ways in which these may be modified by learning, by instruction, and by task design.

The Center issues two series of reports. A Technical Report series includes original reports of experimental or theoretical studies, and integrative reviews of the scientific literature. A Memorandum Report series includes printed versions of papers presented orally at scientific or professional meetings or symposia, methodological notes and documentary materials, apparatus notes, and exploratory studies.

ACCESSION for
CFSTI WHITE SECTION
DDC BUFF SECTION
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
DIST. AVAIL. and/or SPECIAL
A

MISSING PAGE NUMBERS ARE BLANK AND WERE NOT FILMED

T H E U N I V E R S I T Y O F M I C H I G A N

COLLEGE OF LITERATURE, SCIENCE AND THE ARTS

DEPARTMENT OF PSYCHOLOGY

THEORETICAL IMPLICATIONS OF PROACTIVE INTERFERENCE IN SHORT-TERM MEMORY

Raymond Walden Bennett

HUMAN PERFORMANCE CENTER--TECHNICAL REPORT NO. 33

August, 1971

This research was supported by the Advanced Research Projects Agency, Department of Defense, and monitored by the Air Force Office of Scientific Research, under Contract No. AF 49(638)-1736 with the Human Performance Center, Department of Psychology, University of Michigan.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved for public release;
distribution unlimited.

PREFACE

This report is an independent contribution to the program of research of the Human Performance Center, Department of Psychology, on human information processing and retrieval, supported by the Advanced Research Projects Agency, Behavioral Sciences, Command and Control Research under Order No. 461, Amendments 3 and 5, and monitored by the Behavioral Sciences Division, Air Force Office of Scientific Research, under Contract No. AF 49(638)-1736.

This report was also a dissertation submitted by the author in partial fulfillment of the degree of Doctor of Philosophy (Psychology) in the University of Michigan, 1971. The doctoral dissertation committee was: Drs. A. W. Melton, Chairman, R. A. Bjork, W. M. Kincaid, and E. J. Martin.

TABLE OF CONTENTS

		Page
PREFACE		iii
ABSTRACT		vii
CHAPTER		
I.	INTRODUCTION	1
II.	THE EXPERIMENT	23
	Predictions from Theoretical Models	23
	Method	26
	Procedure	26
	Apparatus	27
	Stimulus Materials	28
	Subjects	28
	Design	28
III.	RESULTS	31
IV.	DISCUSSION	43
	Control Processes and Structural Features	43
	Models of Trace Confusability	48
	Classical Strength Theory	48
	An Overview	55
	A Simple Two-State Model	63
	Summary	77
APPENDICES		79
REFERENCES		89

ABSTRACT

Most current models of human memory ignore proactive interference (PI). While this may be a justifiable omission for many tasks, it is certainly true that PI is a major cause of forgetting in situations that involve temporally massed trials with retention measured after an interval filled with a rehearsal preventive activity. It might be imagined that a relatively minor modification would allow some "steady-state" models to handle PI. Typically, however, models that ignore PI make no provision for the possibility that a cause of forgetting is confusion among traces. The primary goal of this dissertation is to demonstrate that such confusions do occur, and that they are of the type that would be expected if such confusions are a primary determinant of PI.

The experiment used a modification of the standard Brown-Peterson technique. The major alteration was to replace the recall test with a two-alternative forced choice recognition test. One of the recognition alternatives (the target) was one of the elements of the current to-be-remembered item (TBRI) and the other (the foil) was an element from some prior TBRI, or it was a word not previously presented in the experiment. The major experimental manipulations were the recency of the foil and the length of the retention interval. Recognition accuracy was found to decrease as the recency of the foil increased and as the retention interval was lengthened. It is argued that this observation is sufficient to make implausible any model not assuming confusion among traces.

An attempt is made to provide a very general structure that appears to contain as special cases most of the current models that rely upon the confusion of traces as a cause of forgetting. This structure is then specialized to produce a simple model that assumes that traces may have only two states, "recent" or "old," and that a response is made by selecting the trace which appears most recent, or by guessing randomly if both traces have the same apparent age (both "recent" or both "old"). The model provides an adequate description of performance on tests using intra-experimental foils, but it breaks down when the foil is a word new to the experiment. Suggestions are made about possible ways of extending the model to make it a more adequate and comprehensive structure, so that it is able to handle not only the recognition data of the current experiment but also the results of conventional recall experiments.

CHAPTER I

INTRODUCTION

Most adults find it easy to multiply two 5-digit numbers together, so long as they are provided with paper and pencil. If deprived of these external aids the same task is quite difficult. It is clear that in mental arithmetic the major difficulty is in the large demands placed upon short-term memory. The same is true for many common information processing tasks. Performance of such tasks will often depend upon the ability to remember relatively small amounts of material for relatively brief intervals during which other activities prohibit active rehearsal. The purpose of this dissertation is to develop some restrictions on the types of models which might be used to represent human memorial capacities under these conditions.

The Brown-Peterson (B-P) short-term memory procedure is of particular interest because it mimes the memory demands of tasks like mental arithmetic. In a B-P experiment Ss are typically presented with a 2- or 3-element to-be-remembered item (TBRI) and then asked to devote their attention to some distracting activity before attempting recall. Devising a suitable quantitative description of forgetting in this situation is of interest not only to those primarily concerned with short-term memory, but also to investigators concerned with other tasks in which memory demands are crucial.

At first glance, the time course of forgetting in a B-P experiment seems easy to describe. For instance, in the Petersons' classic study

(Peterson & Peterson, 1959) the retention function had a form associated with an exponential decay process. The observed values for percent completely correct recall of 3-consonant trigrams and the predicted function from a model postulating a geometric loss are displayed in Figure 1. The model assumes that a trace can be in one of two states, a learned

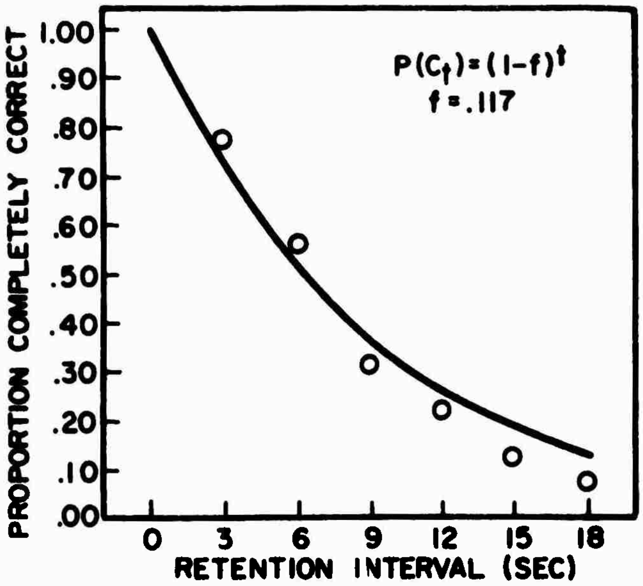

Fig. 1. Data from Peterson and Peterson (1959, Table 1) fitted with a simple geometric loss process.

state which produces perfect recall, or a forgotten state in which recall will be at a guessing level. At the time of presentation an internal representation of the TBRI is created and this representation enters the learned state. During any interval of time, $\underline{dt}$, there is some constant probability, $\underline{f}$, of a trace going from the learned to the forgotten state. If a negligible guessing probability is assumed the probability of recall at a time $\underline{t}$ seconds after presentation will be $(1-\underline{f})^t$.

However, retention functions are often not of a simple exponential form; they typically tend more toward an "S-shape" with an early region of low acceleration. Even these slightly more complicated functions are not difficult to fit with tractable all-or-none forgetting mechanisms. Consider the "dual-trace" model which is presented in Figure 2. It is

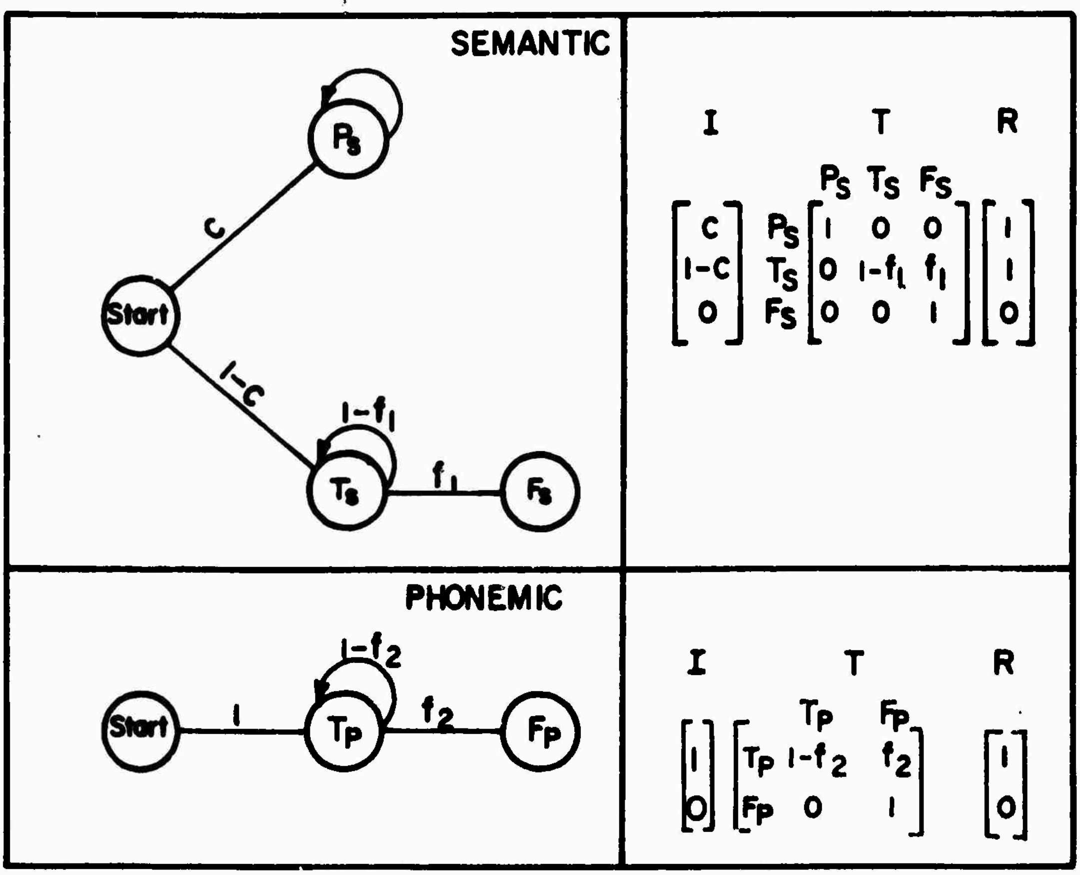

Fig. 2. The left panel shows the state diagram for the dual trace model. The right panel gives the corresponding start vector (I), transition matrix (T) and recall vector (R).

assumed that at presentation two separate traces are constructed (hence the name "dual-trace"), which for convenience have been labelled phonemic and semantic traces. The phonemic trace is transitory and undergoes

the same sort of geometric loss postulated for the Petersons' data. The semantic trace can be of two types. If a "good code" is devised (which happens with probability $\underline{c}$) it is assumed that there will be no forgetting. However, if only a relatively poor encoding is achieved, it is assumed that the trace loss will again be governed by a geometric process. Recall will be successful if either a phonemic or a semantic representation is available; that is, the probability of correct recall will be $p(\underline{S}_t) + p(\underline{P}_t) - [p(\underline{S}_t)p(\underline{P}_t)]$; where $p(\underline{S}_t)$ and $p(\underline{P}_t)$ are, respectively, the probabilities of having a semantic and a phonemic representation available $\underline{t}$ seconds after presentation.

The names semantic and phonemic should not be taken too seriously. First, the reasoning behind the model was that one kind of information is at least potentially permanent while the other kind is necessarily transitory. The terms semantic and phonemic were selected only because they seemed to fit with the notions about the nature of the internal representation that have developed in conjunction with multi-store models. Second, the formal structure of the model makes it difficult to preserve a rigorous distinction between forgetting attributable to loss of phonemic information versus degradation of semantic information. The probability of a correct recall can be expressed as:

$$p(\underline{C}_t) = \underline{c} + (1-\underline{c})(1-\underline{f}_1)^t + (1-\underline{f}_2)^t - (1-\underline{f}_2)^t[\underline{c}+(1-\underline{c})(1-\underline{f}_1)^t]$$

This reduces to:

$$p(\underline{C}_t) = \underline{c} + (1-\underline{c})[(1-\underline{f}_1)^t + (1-\underline{f}_2)^t - (1-\underline{f}_1)^t(1-\underline{f}_2)^t].$$

When expressed in this way, it is clear that there is a trade-off between the values of $\underline{f}_1$ and $\underline{f}_2$. Any particular theoretic prediction will not be altered if the values of $\underline{f}_1$ and $\underline{f}_2$ are interchanged. Given this fundamental ambiguity, there is, of course, no particular significance in identifying one parameter with phonemic and the other with semantic information.

Figure 3 shows the fit of the dual-trace model to the data from an experiment by Hellyer (1962) in which presentation frequency was

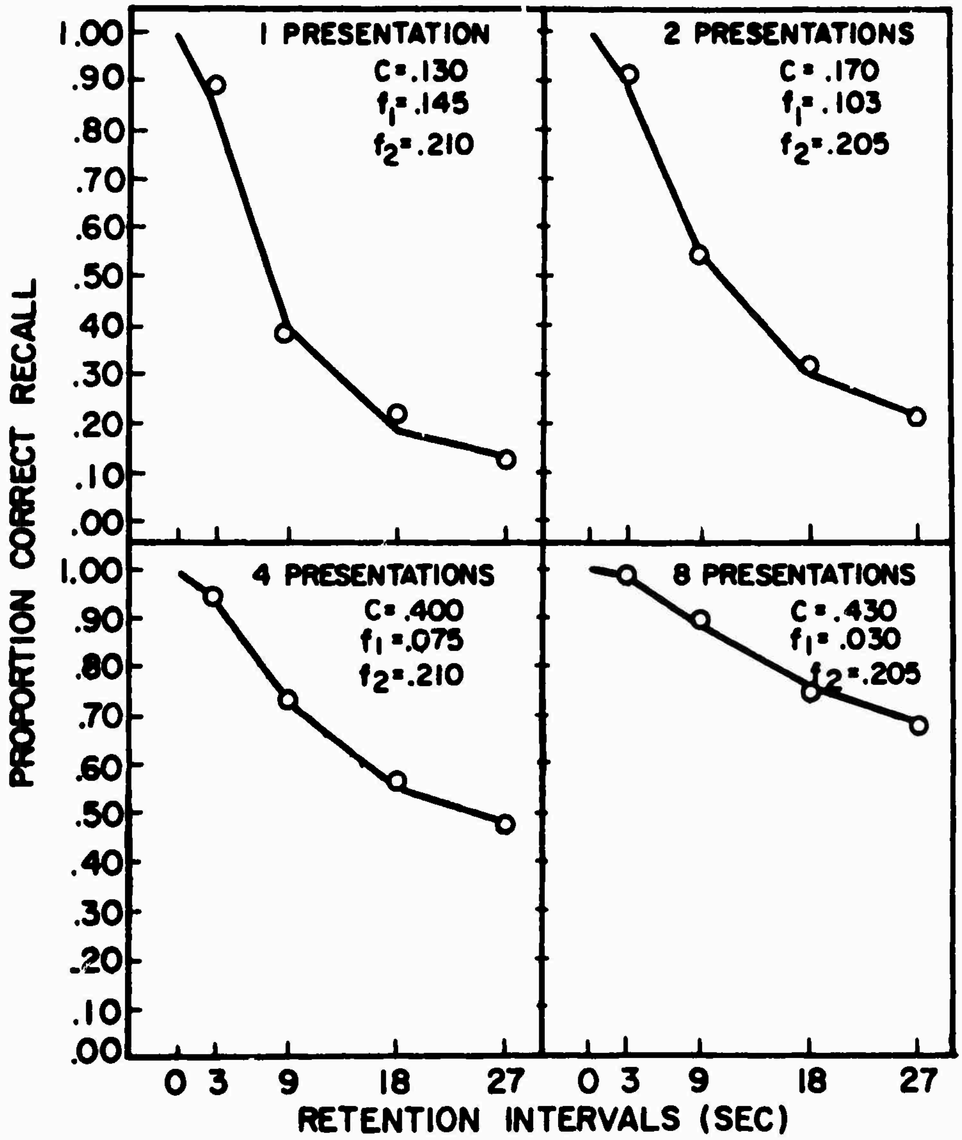

Fig. 3. Fits of the dual-trace model to data from Hellyer's (1962) experiment.

manipulated over four levels (1, 2, 4, or 8 1-sec. presentations of low-M consonant triples) and retention was measured after 3, 9, 18, or 27 sec. As can be seen the fits are quite good and the parameters vary in a lawful way with the presentation frequency manipulation. With increasing numbers of presentations the probability of getting a good code goes up, one of the forgetting parameters decreases and there is no apparent change in the other.[1]

In fitting Hellyer's data the dual-trace model demonstrated its capacity to accept any of the retention functions as possible outcomes of a B-P experiment. However, the demonstration of acceptable fits does not explain what repeated presentations accomplish. An adequate accounting of Hellyer's data would have to include an auxiliary model that predicts the observed changes in the parameter values of the dual-trace model.

The author has fitted the dual-trace model to a large number of retention functions from B-P experiments and it appears to be an adequate representation of memorial capacity. It would appear that this model, or something similar to it, is a satisfactory solution to the problem with which this dissertation began. Unfortunately, some complexities have been ignored. Most notably, recall in the B-P situation is subject to strong proactive interference (PI). Keppel and Underwood (1962) performed a set of experiments in which

---

[1]Since there are only 4 data points in each retention function, and since the model has 3 free parameters, these data could not be used as a test of the model. However, the fit is not materially damaged if one forgetting parameter, $\underline{f}_2$, is held constant across all presentation conditions. This results in a total of 15 dfs in the data and 9 parameters.

they varied both the retention interval (3 and 18 sec.) and the number of prior trials (0 to 5). On the first trial retention is not noticeably affected by delaying recall, but on succeeding trials the retention function becomes progressively steeper. The data from their third experiment are plotted in Figure 4.

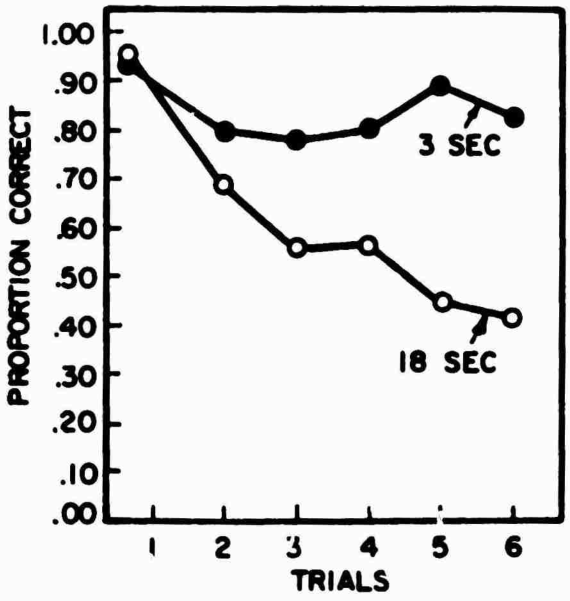

Fig. 4. Proportion of correct responses as a function of number of trials and retention interval. (Data from Keppel & Underwood, 1962, Exp. III, Figure 4)

For the dual-trace model the formal status of PI is very similar to that of the frequency of presentation manipulation used by Hellyer. In order to claim an explanation of PI it must first be shown that the dual-trace model can fit retention functions under any specific level of PI. Then it is necessary to construct a second model which will explain why increasing the number of prior trials induces the observed changes in the parameter values of the dual-trace model.

The first step, applying the model to retention functions obtained under all levels of PI, is easily accomplished. While there are few studies which provide sufficient degrees of freedom to test the model, it has been successfully applied to several published and unpublished studies. The most comprehensive investigation of the build-up of PI is a study by Noyd (1965). Noyd presented 2-, 3-, and 5-word TBRIs, and measured retention after 4, 8, and 24 sec. The obtained PI build-up functions are displayed in Figure 5. There is a decrement in performance over the

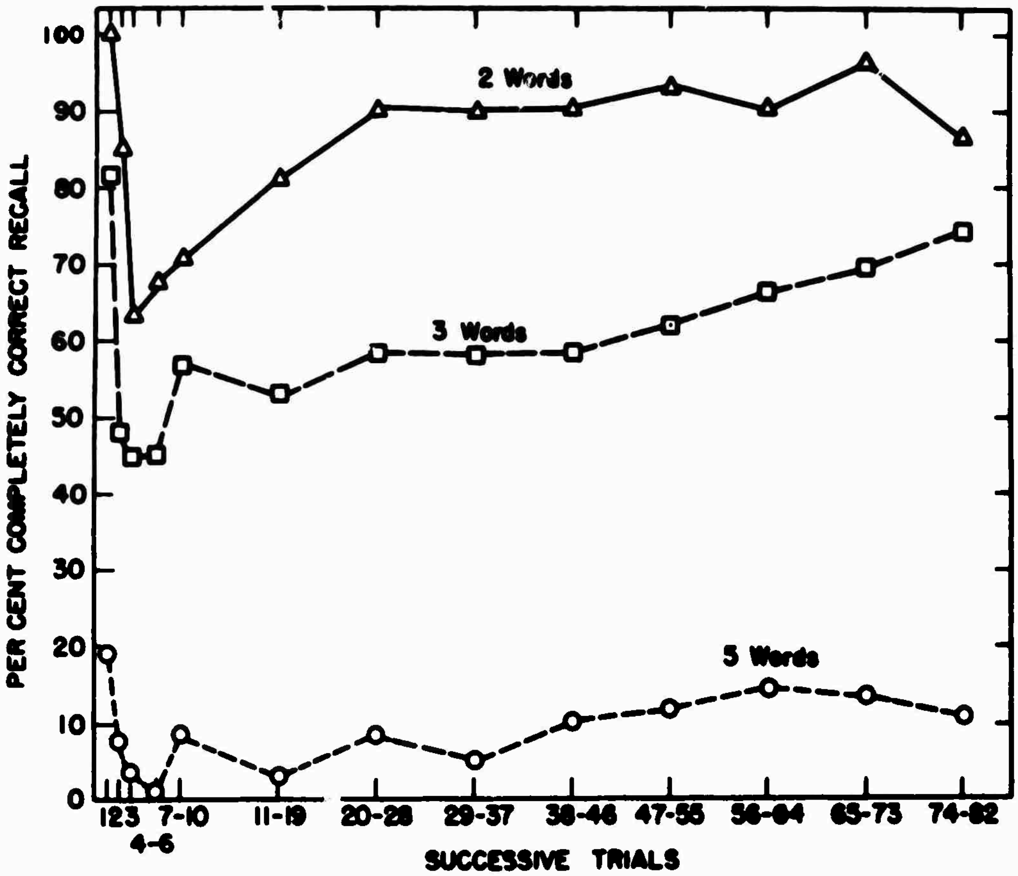

Fig. 5. Retention as a function of number of trials and length of TBRI. (Data from Noyd, 1965)

first few trials followed by a gradual improvement lasting for several tens of trials. This is a typical result with naive Ss and undoubtedly reflects a practice effect. The best data for testing the dual-trace model are for the first few trials of the 3-word condition. The observed values and best-fit retention functions are shown in Figure 6.

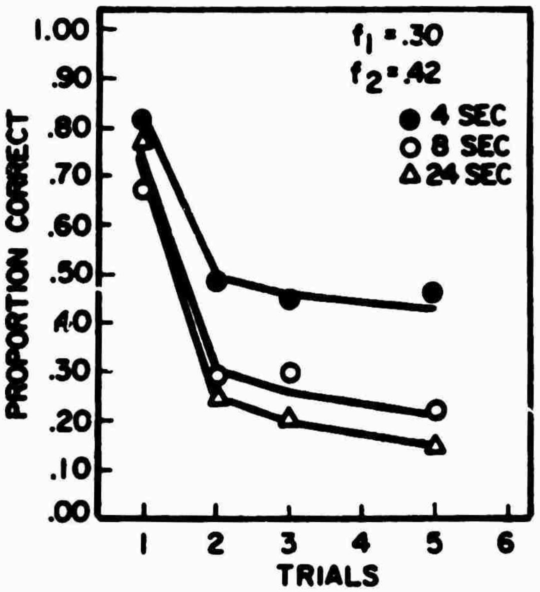

Fig. 6. Fits of the dual-trace model to PI build-up functions. The forgetting parameters (f_1 and f_2) were held constant and c allowed to vary. The values of c, for Trials 1 to 4-5-6 were, respectively, .72, .25, .20, .15. (Data from Noyd, 1965)

The fits in this figure were made by allowing c (the encoding parameter) to vary freely while holding f_1 and f_2 constant for all of the functions. The obtained fits are about as good as could be expected for data as variable as these (there are only 27 observations per data point). Fits were also attempted by allowing c to covary with f_1 or f_2 but there was no noticeable improvement. It was not possible to obtain reasonable fits when c was held constant across all levels of prior trials.

This result, if it is generalizable to other B-P experiments, demonstrates the feasibility (actually, it does not deny the feasibility) of obtaining an account of PI within the framework of the dual-trace model. The obtained pattern of changes in parameter values is suggestive of the sorts of models which might be able to handle PI. Of most importance it would appear that PI acts by controlling the utilization of semantic (i.e., potentially permanent) information. Specifically, there is the hint that increasing numbers of prior trials decreases the probability that Ss will happen upon an encoding which will be adequate for developing a permanent representation of the information. This might mean that the efficiency of encoding was being decreased through the action of some fatigue-like process. However, it is just as reasonable to assume that the quality of the codes generated by Ss undergoes no decrement, but that the requirement for an adequate code becomes more stringent with increasing numbers of prior trials.

While it might be possible to develop an adequate account of PI using the same general structure as the dual-trace model, there are some data which at first seem to deny the feasibility of such an approach. Of particular interest are the kinds of errors that Ss make. It is possible to divide responses into four unambiguous classes. A correct response is credited whenever all of the elements of the TBRI are reported in the correct order. A transposition error is said to occur whenever the reported elements of the TBRI are in the wrong relative order. A response is called an intrusion when

S reports out a possible response which was not an element of the TBRI. An intrusion may be all or part of some prior item (an intra-experimental intrusion) or it may be an importation from outside the experiment (an extra-experimental intrusion). Finally, Ss occasionally opt to make no response at all. and an omission is scored.

The most interesting of the errors are the intra-experimental intrusions. Since the theoretical status of intrusions will be an important point, a more thorough documentation of the literature is desirable. The source of an intra-experimental intrusion will be defined as the TBRI in which the intruding element was originally presented, and the source lag (or lag) as the number of trials intervening between the source and the current trial (counting the current trial). (It should be noted that the source of an intrusion may be defired unambiguously only when the elements of a TBRI are unique to that item. This usually means using word TRBIs.) It is then easy to describe one of the most powerful determinants of intrusion frequency: The probability of an intrusion is inversely proportional to the source lag. This relationship is illustrated in the data (from Noyd, 1965) plotted in Figure 7. The proportion of intrusions decreases regularly with source lag, reaching a near 0 value when the source lag is about nine items. (The "correction for opportunity" used in plotting Figure 7 takes into account the fact that Ss have a greater opportunity to make intrusions from more recent sources. For instance, in a 10 item sequence there are 9 opportunities to make intrusions with Lag 1, but there is only one trial (the last) on which an intrusion

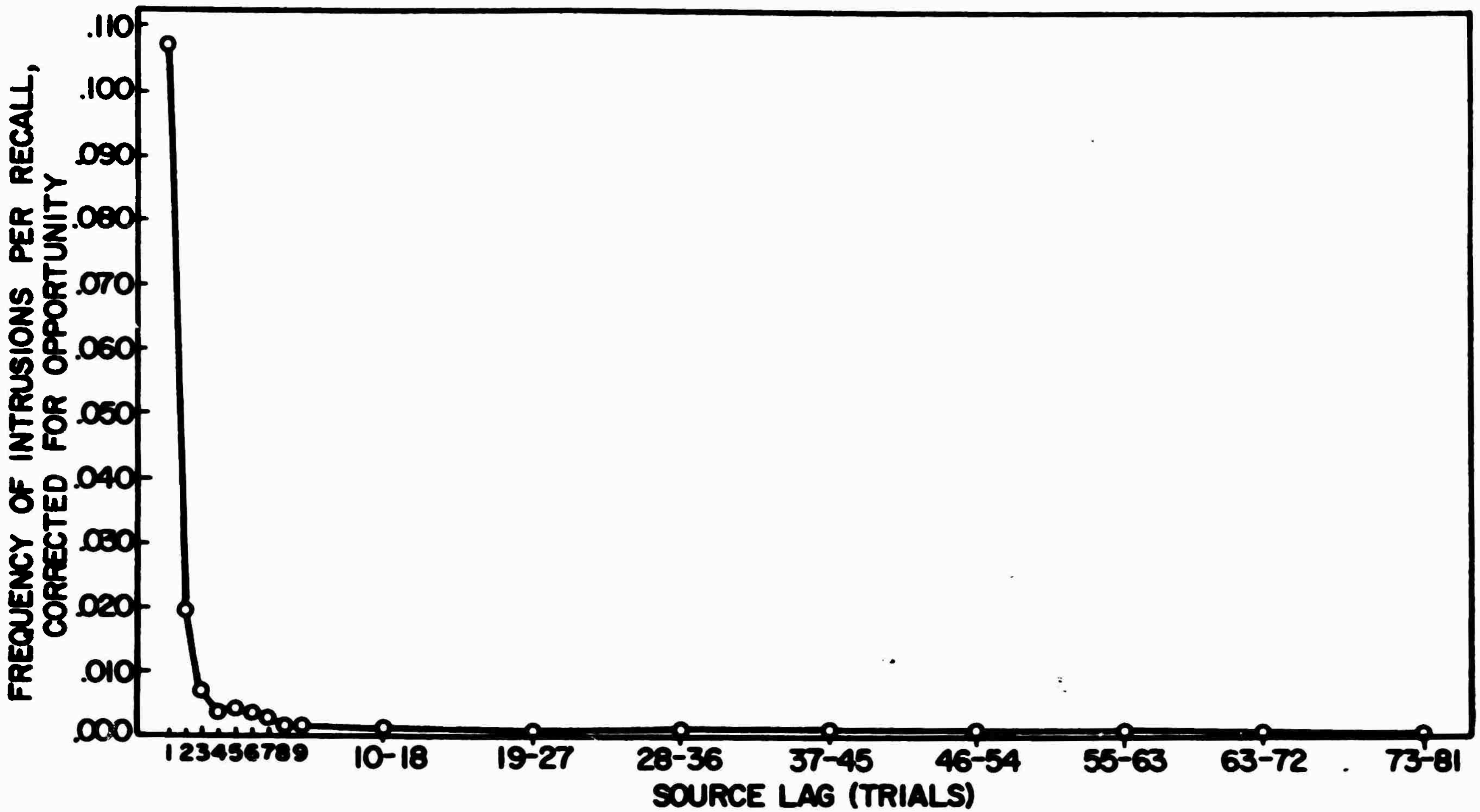

Fig. 7. Intrusion rate as a function of source lag. (Data from Noyd, 1965)

of Lag 10 can be produced. Noyd used an 82-trial sequence. For the data plotted in Figure 7 the correction factor is large only for the most extreme lags.)

There is a strong tendency for intrusions to come from sources which are similar to the item presented on the current trial. For instance, in a set of experiments by Loess (1967, 1968) TBRIs were drawn from different natural language categories. There was a powerful tendency for intrusions to be from the category that was appropriate to that trial. Whenever an intrusion was from the wrong category, it was still the case that all the elements produced on any particular trial came from the same category (Loess, personal communication).

An obvious formal similarity is the serial position of an intruding element within the source and its serial position as an intrusion. When attention is restricted to intrusions all of which are of Lag 1 it is not uncommon to find up to 80% of the intruding elements being in the same within-item serial position as they had occupied in their original presentation. This relationship is illustrated in the data presented in Table 1, which comes from an experiment by Fuchs and Melton (unpublished). The left panel is for 3-word TBRIs, the right panel for 5-word TBRIs. The decrease in serial position specificity of intrusions with increasing TBRI length is typical in B-P studies that require ordered recall.

In Noyd's (1965) experiment <u>S</u>s saw either 2-, 3-, or 5-word TBRIs. Consequently, it is possible to look at the frequency of intrusions as a function of the commonality of the length of the source and the TBRI. The pertinent data are presented in Figure 8. Of most

TABLE 1

SERIAL POSITION OF OCCURRENCE OF AN INTRA-EXPERIMENTAL INTRUSION (FIRST OCCURRENCE ONLY) AND ITS SERIAL POSITION IN A PRECEDING SOURCE STIMULUS

(Fuchs & Melton, Unpublished)

Serial Position in Source	3-Word TBRIs Serial Position of Occurrence in Recall 1	 2	 3	Total
1	29	9	4	42
2	8	32	5	45
3	3	7	36	46
TOTAL	40	48	45	133

	5-Word TBRIs 1	 2	 3	 4	 5	
1	17	23	8	15	6	69
2	10	19	18	19	9	75
3	10	13	30	26	4	83
4	11	18	19	32	8	88
5	9	13	30	17	29	98
TOTAL	57	86	105	109	56	413

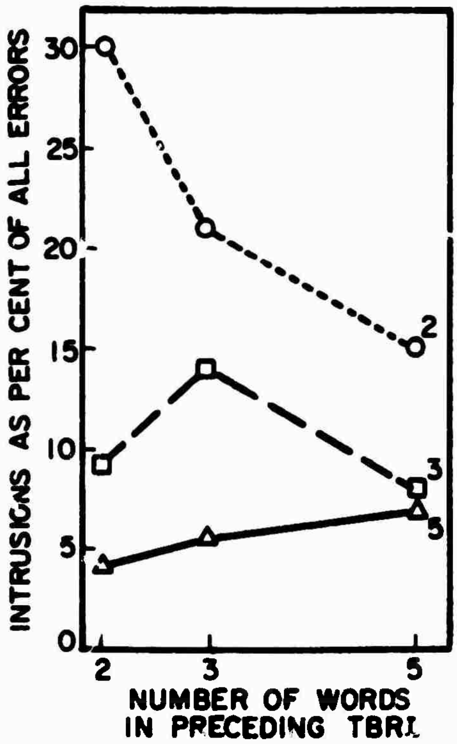

Fig. 8. Intrusion rate from preceding 2-, 3-, and 5-word TBRIs as a function of TBRI length. (Data from Noyd, 1965)

interest is the non-monotonic appearance of the function for 3-word TBRIs. It is quite apparent that intrusion frequency is maximized when the TBRI and the source are of the same length.

Finally, there is a strong suggestion that the frequency of intrusions and the probability of a correct response are inversely proportional. Consider the data plotted in Figure 9 (from Noyd, 1965). The solid line is the probability of a correct response as a function of ordinal position of a trial in the experiment and the dotted line is the probability of an intrusion. Plotted in this way, it is obvious that the two functions are mirror images.

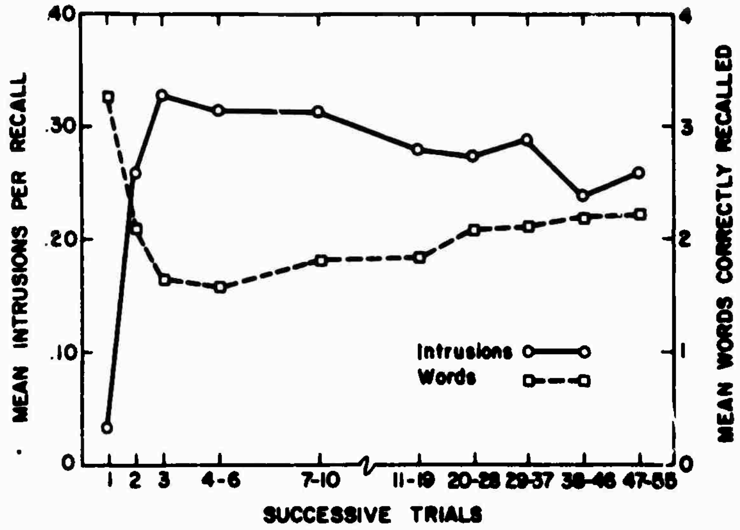

Fig. 9. Retention and intrusion rate as a function of number of trials. (Data from Noyd, 1965)

The data on intrusions appear to represent a serious challenge to the dual-trace model. As it stands there is no provision for the occurrence of intrusions. A trace either is or is not available. If it is available, recall will be correct; if it is not, the system is in a state where no response can be made. Fortunately (or not), it is not particularly difficult to revise the model so it can account for at least some of the intrusion data. For instance, it is possible to obtain a reasonably accurate quantitative description of the relationship between source lag and intrusion rate. The model predicts that at any time in a sequence of B-P trials there will be a set of traces of prior items which are still in one of the learned states. If the TBRI has entered a forgotten state, then the retrieval system randomly selects one of the available traces as a response. This scheme leads to the predicted function in Figure 10 (the data are the same as in Figure 7).

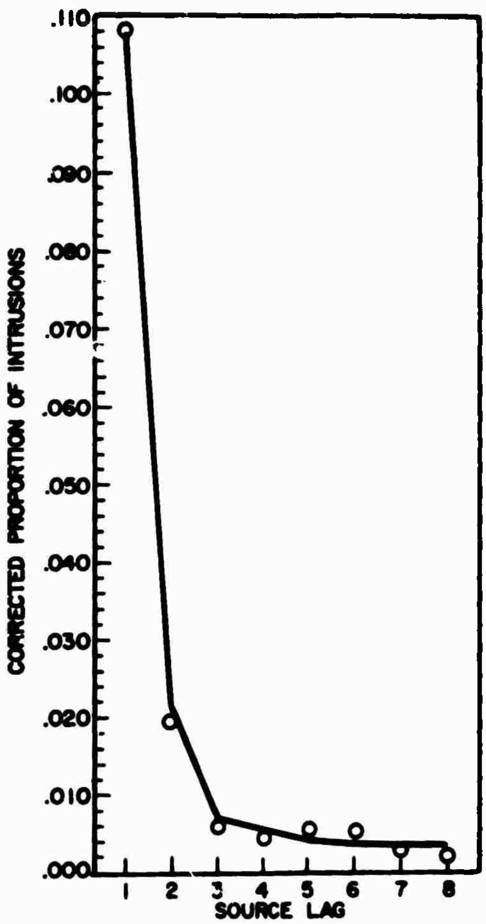

Fig. 10. Observed and predicted source lag functions. Predictions were made from the dual-trace model assuming that when the TBRI was forgotten an intrusion would be produced on a constant proportion of the trials. An intrusion is assumed to be randomly selected from among the traces still in one of the learned states. (Data from Noyd, 1965)

The apparent reciprocity between intrusion rate and probability of recall (see our Figure 9) is equally easy to dispose of. A close examination of the figure reveals that it demonstrates only that intrusions are a more or less constant proportion of the errors. This relationship is handily incorporated into models like the dual-trace model.

It is possible to develop other elaborations of the model that would allow it to handle the rest of the intrusion data. However, if the dual-trace model were accepted as an essentially accurate representation of forgetting, an account of intrusions would be only window

dressing. For the dual-trace model intrusions play no part in the processes responsible for forgetting. They are solely the result of strategies used by the S when a response must be produced in the absence of sufficient information.

This is not the only available conceptualization of intrusions. For several different sorts of models, intrusions are not the result, but rather the cause of errors. Since these formulations represent important alternative conceptualizations of the forgetting process, it is worth while to consider them in more detail.

Conrad (e.g., 1960, 1967) has suggested that the internal representation of the TBRI can be described as a bundle of distinctive features. If sufficient time is allowed for encoding, a newly created trace will unambiguously indicate a single response. However, if recall is delayed, information may be lost from the trace and there will be several possible responses which might be represented by the remains of the trace. If, at the time of recall, the trace has been degraded so that it does not lead to a unique response, a single response can be generated by selecting one possibility from the equivalence class defined by the trace residue.

This conceptualization predicts that intrusions will occur in substantial numbers. Early in the retention interval the equivalence class defined by the trace will be fairly small, and errors will tend to be related to the TBRI. After a longer delay the trace will be more severely degraded and errors will tend toward randomness. Conrad (1967) has obtained data in agreement with this notion. Using consonant

quadrigram TBRIs Conrad found that early in the retention interval the erroneous responses tended to be phonemically related to the TBRI. Errors produced after longer retention intervals were much less systematic.

However, a notion like this faces some difficulties in handling other intrusion data. It is not at all obvious why the source lag of intrusions is predominantly short. The only obvious explanation would be that the trace contained information about the recency of presentation. It would, of course, be difficult to prove that such was not the case. It is unclear, however, what purpose such information would serve. The only information in a trace which would be of apparent use would be whatever was necessary to get back to the original material. Recency information would not appear to be particularly useful for this purpose. Nonetheless, it would likely be possible to construct a model which assumed that traces incorporate information about the time at which they are created, and thus handle the source-lag effect. While such a model might appear *ad hoc*, it would probably be difficult to refute and might well do a reasonable job of handling the data.

Interference theory offers yet another alternative. A possible analysis of the B-P situation would assume that at the time of presentation the trace of the TBRI is associated with the "context." Over time the strength of the association declines. At the time of recall the retrieval system examines all of the traces in store and selects the one that is most strongly associated with the context. In order to predict some reasonable error rate, it might be assumed that the associative connections fluctuate in strength over time, and the recall

decision is made by selecting the trace which is momentarily most strongly associated with the context. (Keppel & Underwood, 1962, have presented a slightly different analysis. The most striking difference between their position and the one suggested here is their application of the notions of "extinction" and "spontaneous recovery." This will be discussed in Chapter IV.)

For a model of this sort, all errors are the result of intrusions. The only time the TBRI is not correctly reported is when S selects the trace of some other item in preference to the TBRI. The immediate difficulty with models of this class is that they predict an intrusion rate that is altogether too high. This is not an insurmountable difficulty. It would be quite easy to assume something analogous to "associative blocking" (e.g., Postman, 1961). For instance, the separation of the momentary trace strengths might have to exceed some critical value before the S would choose to emit a response that might be in error.

A competition model of this sort has no difficulty in accounting for effects like the relationship between intrusion rate and source lag. The major stumbling blocks are encountered when attempts are made to develop descriptions of the mechanisms which are responsible for producing intrusions that are similar to the TBRI. The model, as it has been presented, makes no distinction between traces on the basis of their content. The decision about which trace to output is governed only by the degree to which the traces are associated with the effective recall stimulus. As Conrad (1967) has pointed out, there would be no

reason to expect the trace representing the letter "b" to replace the TBRI "t" more often than the letter "f."

Like the other models considered here the competition model can be handily modified to bring it into a closer match with the data. The most straightforward way to produce a more complete account of the intrusion data would be to expand the kind of information which the retrieval system uses in selecting a response for recall. At the moment, it is assumed that the only criterion used in deciding which response to make is the strength of the association between the context and a trace. One may reasonably assume a retrieval system in which a trace must pass several tests prior to S deciding upon it as the most likely basis for an overt response. For instance, it might be assumed that at the time of presentation the retrieval system informed itself of some of the characteristics of the TBRI of that trial (e.g., the natural language category, the number of words being presented, etc.). At the time of test this information would be used to screen out impermissible responses which might be associated with the context. While a notion like this might become somewhat unwieldly, it is likely that it could be made to predict most or all of the intrusion data.

In summary, there are several competing formulations of forgetting processes all of which appear potentially able to provide a description of proactive effects in the B-P situation. They make very different assertions about the import of intrusions, but it is not clear that any conventional recall experiment could ever provide data which would deny the feasibility of any of these types of models. While

further study of intrusions might be of no particular importance, a resolution of the theoretical issues raised in attempts to make sense of the intrusion data would be of considerable significance. If it could be determined that intrusions were just futile guesses it would be possible to discard any model like Conrad's or the competition model. Alternatively, if intrusions proved to be a necessary concomitant of the forgetting process, models like the dual-trace model would be falsified. One of the main goals of the experiment proposed in the next chapter is to provide definitive data on this matter.

CHAPTER II

THE EXPERIMENT

A slight modification of the B-P procedure makes it possible to solve problems that are intractable with the standard procedures. The modification suggested here replaces the usual recall measure with a 2-alternative forced choice (2-AFC) recognition test in which one of the alternatives (the target) is an element of the TBRI and the other (the foil) is an element from some past TBRI or from outside the experiment. The major experimental manipulation is the foil lag, which is the number of trials, counting the current one, since the original presentation of the foil. The value of this technique is most easily seen when the predictions from the three models are compared.

Predictions from Theoretical Models

Conrad's trace-degradation model predicts that the nature of the foil will be a powerful determinant of the probability of a correct response. The most straightforward analysis assumes that S first generates internal representations of the two alternatives. The alternative whose representation affords the best match to the trace of the TBRI is selected as the correct response. The original formulation of the trace degradation model predicts that foil lag is unimportant; the only interesting characteristic of the foil is its similarity to the TBRI. However, as was seen in the discussion of intrusions, Conrad's statement of the model will have to be expanded to include within a trace some representation of the "recency" of presentation.

With this addendum, the structure predicts that short-foil lags will produce worse performance than foils of longer lags. The best performance is expected when the foil is a "new" word.

The competition model makes very similar predictions. It is assumed that the strength of each alternative is assessed, and the strongest selected as the response. This would produce best performance when the foil is weakest, so recognition would again be expected to increase with increasing foil lag. It should be noted that this is identical to attempting recall when there are only two traces in store. In fact, for both the trace-degradation model and the competition model, the processes which lead to intrusions are identical to those responsible for errors in the recognition test. As will be seen in Chapter IV, the recognition stiuation is considerably more tractable than is recall for analyzing models of this sort.

An analysis based on the dual-trace model starts with the assumption that if the trace of the TBRI is still in store at the time of test, the correct alternative is always selected. There are several alternative schemes which might be used to generate responses when the trace has entered the forgotten state. The simplest would be to pick one alternative at random. This leads to the prediction that foil lag is ineffective. However, it is possible to postulate more complex response strategies. One of the most efficient relies on the fact that if the trace of the TBRI is not available, S knows that any trace still in a learned state will be from a prior trial. Consequently, S can adopt the strategy of examining the store to

see if there are internal representations corresponding to either alternative. If such a representation is located it is highly likely that the alternative it represents is from a prior trial, and S can, by elimination, assume that the other alternative is the correct one. This scheme results in most accurate recognition of the TBRI with the shortest foil-lag and a steady decrement in performance with increasing lag, until a minimum is reached for cases where the foil is "new."

Of most immediate interest, it does not seem to be possible to make the dual-trace model, or any model "similar" to it, predict increasing recognition performance with increasing foil lag. This is easily seen if an attempt is made to discover the conditions under which the dual-trace model will make this prediction. It was assumed that recognition performance will be perfect when the trace of the TBRI is in one of the remember states. When the trace of both the foil and the TBRI are unavailable there is no information upon which to base a response, so it must be assumed that performance will be at a chance level. Consequently, at an arbitrarily long lag, the probability of a correct recognition will be $p + 1/2(1-p)$, where p is the probability of the TBRI still being available. If recognition is to get worse as the foil lag is shortened it must be the case that the guessing strategies are driving performance <u>below</u> chance levels. While it is certainly possible to postulate models with this property, it is doubtful that it would be appropriate to call them guessing models. In fact, it such a mechanism were superimposed on the dual-trace model, it would become a competition model, and act just like a system subject to confusion about trace identity.

In summary, if the foil-lag effect is as predicted by the competition and trace-degradation model, then it seems safe to deny the feasibility of models like the dual-trace model. Alternatively, if there is no foil-lag effect, or if performance decreases with increasing lag, then there is reason to doubt the reasonableness of the competition analysis afforded by interference theory, and that of Conrad's trace-degradation model.

Since the arguments used in arriving at this conclusion depend upon only a few characteristics of the presented models, it is likely that this experiment will afford conclusions of greater generality than those claimed here. This will be considered in more detail in Chapter IV.

Method

Procedure

The basic procedure is a modification of the standard Brown-Peterson technique. Every trial began with the word "READY" which was on for 2 sec. Immediately thereafter three 4-letter nouns were presented sequentially for 500 msec each. The S read the words aloud, and then worked on a serial reaction time task for 8, 20, or 30 sec. During this interval the S was required to call out the name of the first presented number and press one button for an odd number and the other button for an even one (the even button was always under the first finger of the dominant hand). As soon as a button was depressed a new number would appear and the S again named it and made

a keypress response. It was possible to earn up to $1.00 in bonus pay by responding quickly (within 500 msec) to the first number that came up and by maintaining a high rate of accurate responding (1.8 digits/sec at 92% accuracy) during the entire retention interval. There was no payoff for accurate memory performance and Ss were encouraged not to rehearse the TBRI during the retention interval. At the end of the retention interval the word "TEST" appeared for 7.5 sec. The S turned over the top card on a deck of cards, circled a number on his answer sheet, and then placed the card face down. On each card there were two different 4-letter nouns, a correct alternative from the TBRI of that trial and a foil. The S decided whether the right or left word was from the TBRI and marked his answer sheet accordingly. On each line of the answer sheet there were 8 numbers arranged in this manner:

4 3 2 1 1 2 3 4

The left numbers were to be used if the left word was the correct choice and the right numbers for the right word. The number "4" indicated high confidence and "1" was used for guessing level confidence. Left and right responses were equally often correct. At the end of the test, the word "REST" was presented for 3 sec and then the onset of a new trial was signaled with the word "READY."

Apparatus

All stimulus events (with the exception of the alternatives used for the recognition test) were presented on a CRT controlled by a PDP-1 computer. The S sat in a moderately sound-attenuating cubicle with two fans that provided some auditory masking.

Stimulus Materials

All words were common single-syllable 4-letter nouns with a Thorndike-Lorge (1944) word count of at least 10 per million. The three words used to make up any particular item were selected so that no two of them shared a common first letter and so that there were no obvious formal, semantic, or phonemic relationships.

Subjects

The Ss were 108 male and female undergraduates at the University of Michigan who were attending the summer session or working in Ann Arbor between semesters. They were paid $2.00 (with a maximum bonus of $1.00) for participating in the 1.25-hr session.

Design

The experiment was a 6 (foil lag) x 3 (retention interval) x 3 (within-item serial position of the correct alternative) x 3 (foil within-item serial position) factorial. The foil lags were 1, 2, 4, 8, 12, and "new," and the retention intervals were 8, 20, and 30 sec. Since there were 162 conditions it was not feasible to have a completely within-S design. However, collapsing over the within-item serial positions, each S saw 3 replications of the complete foil lag by retention interval (6 x 3) design; collapsing over retention interval there was one replication of the foil lag by within-item serial position design (6 x 3 x 3); and collapsing over foil lag there were 2 replications of the within-item serial position by retention interval (3 x 3 x 3) design.

Each S saw a series of 61 trials, 7 starters[2] which were were not scored, followed by 54 experimental trials. In order to scramble the order of foil lag conditions and to counterbalance items and retention intervals, 36 separate sequences of conditions were generated. All 54 TBRIs were segregated into 18 sets of 6. The TBRIs within a set of 6 were rotated in accordance with a balanced Latin square, producing 6 orders of 54 TBRIs. In all 6 orders it was always the case that no two words on adjacent trials in the same within-item serial position shared a common first letter or had any strong semantic or phonemic similarities. Each order of the TBRIs was assigned a single sequence of conditions. The major constraints in constructing these sequences were that (a) every foil type and retention interval would occur equally often in every third of the experiment, and (b) no word would be used more than once as a recognition alternative. Six new orders of conditions were generated from each of the original 6 by rotating the values of the retention intervals. This procedure insures that each retention interval will appear equally often with every other condition and that within the 36 sequences the full 6x3x3x3 design will be replicated 12 times. The sequential effects were not completely controlled, but, over Ss, each retention interval was preceded and followed equally often by each retention interval, including itself,

---

[2]This was an error. The original design called for 13 starters, which insures that every lag can be used in every one of the 54 experimental trials. By chance there were no long lag tests near the beginning of any of the sequences and the additional 6 starter trials were accidentally omitted.

and within each S the first-order sequential effects were scrambled (the ordering of retention intervals within any one of the 36 sequences was almost random).

CHAPTER III

RESULTS

Foil lag had an orderly effect on recognition performance. In Figure 11 the probability of an error is plotted as a function of foil lag. Error rate is a monotonically decreasing function of foil lag, with performance at the longest lag reliably worse ($p < .005$) than when the foil is a word new to the experiment. There is better than a 3-fold increase in error rate going from "new" foils to foils drawn from the immediately preceding trial.

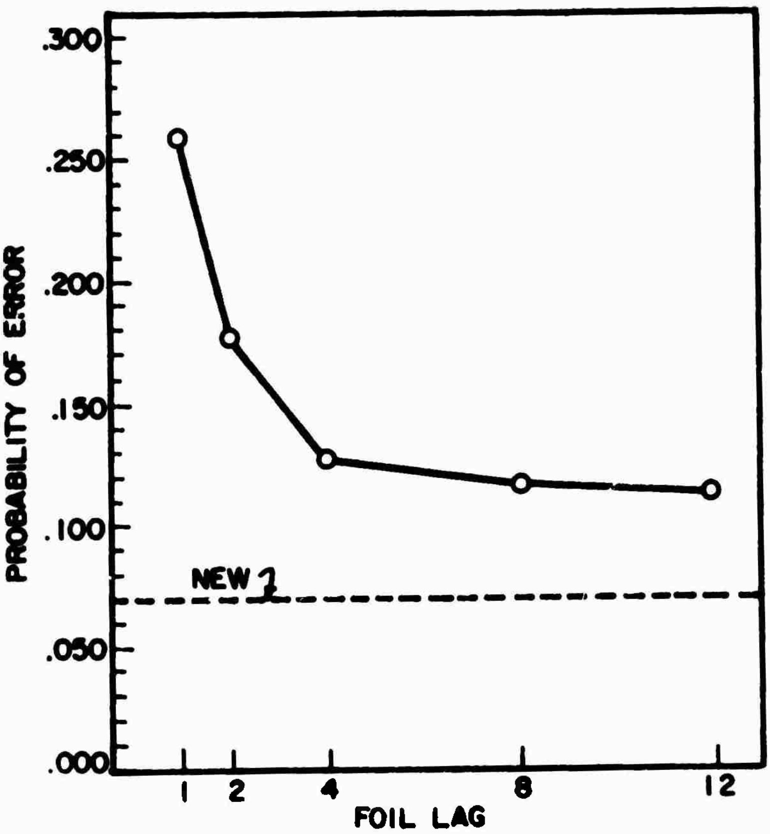

Fig. 11. Probability of incorrectly accepting a foil as a function of foil lag. There are 972 observations per data point.

Since confidence ratings were used, it is possible to plot receiver operating characteristics (ROCs, Green & Swets, 1966) which show the trade-off between hit and false-alarm rates. This has been done in Figure 12. The effect of foil lag is again apparent, although, with this measure, the difference between foils of Lags 4 and 8 is less than it is in the error rate data. The area under the ROC curve (A_G, Green & Swets, 1966) was computed for each foil lag and the values are plotted in Figure 13. Quite clearly the foil lag effect has the general form expected by the competition models. The data appear to be in direct contradiction to the predictions of models like the dual-trace model.

An interesting incidental question is whether the probability of correctly rejecting a foil is related to the probability that the target from the same word triple was correctly accepted. If it were reasonable to believe that Ss were handling the word triples as a unit or "chunk" then it might be expected that the forgetting of a TBRI would be "all-or-none" (cf., Johnson, 1970). This in turn would lead to the expectation that the probability of correctly rejecting a foil would be dependent upon whether the target from the same item had been correctly accepted. (The direction of this dependancy is not obvious. It would probably be possible to develop reasonable *post hoc* explanations for effects in either direction.) In these data there is no sign of such a dependency. Averaging over all foil lags, the probability of a correct rejection of the foil, given that the target from that item was accepted as correct (.1470) is almost identical

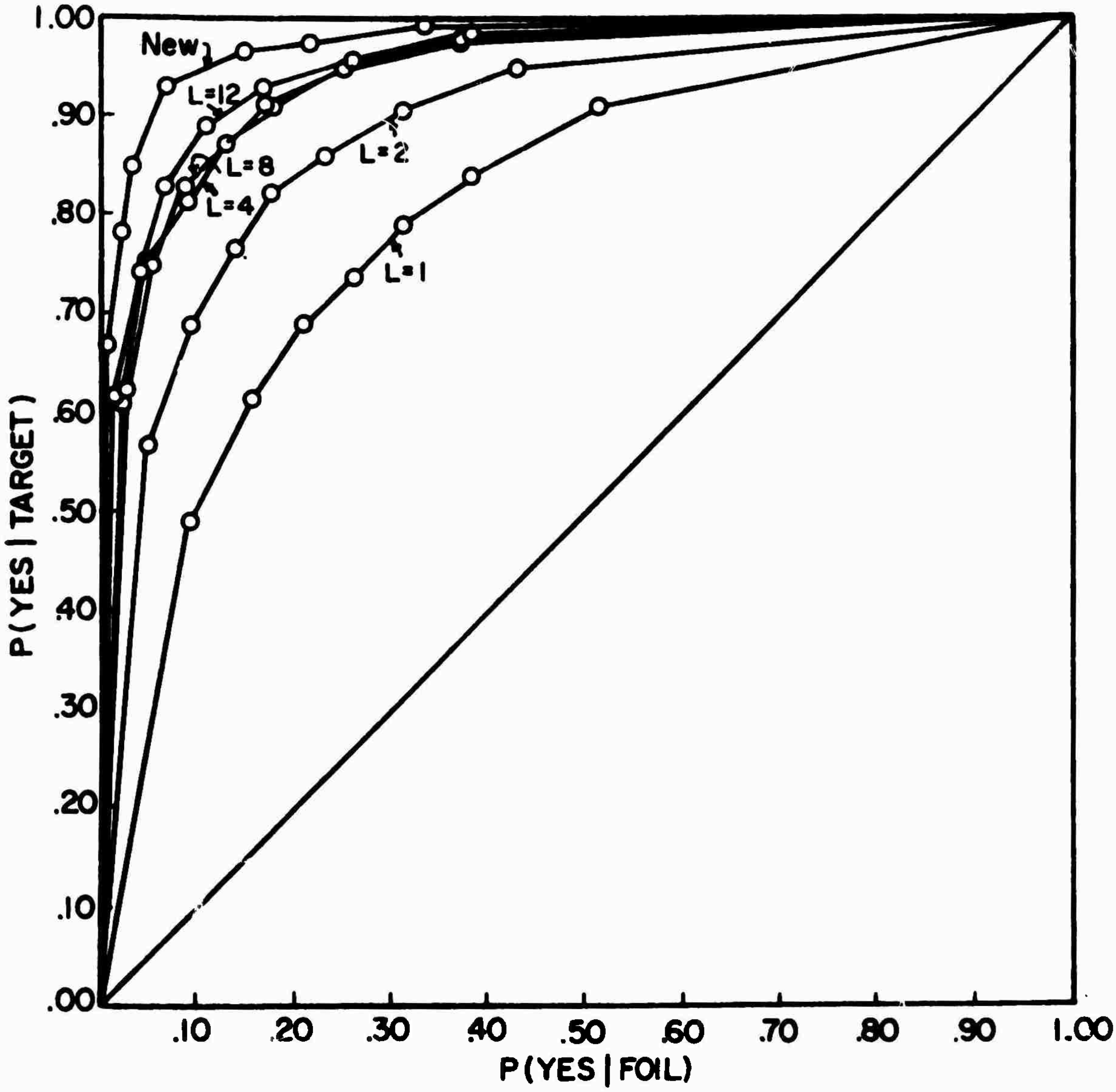

Fig. 12. ROCs for the different foil lag conditions. There are 972 observations per curve.

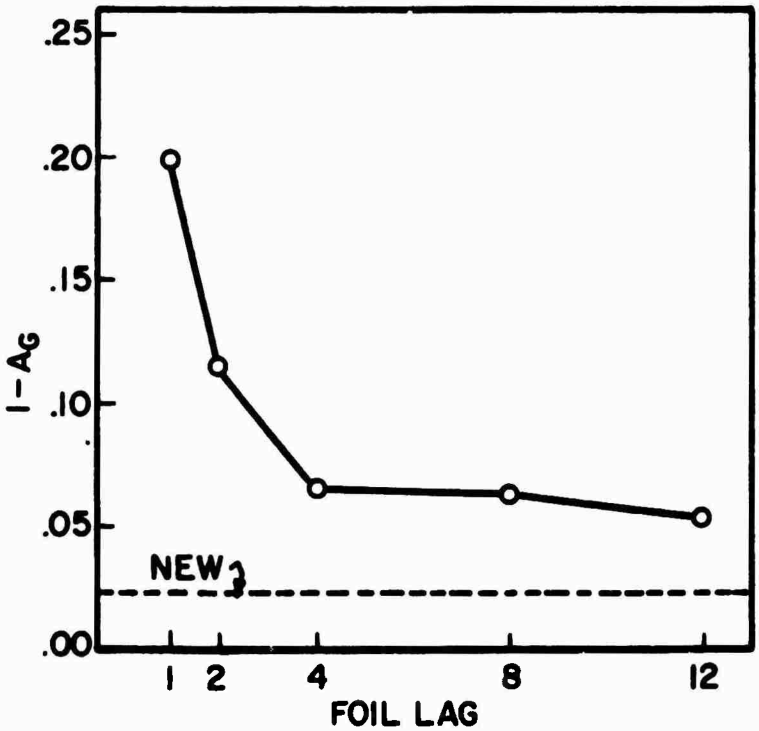

Fig. 13. Probability of error ($1-A_G$) as a function of foil lag.

to the probability of a correct rejection of a foil when the corresponding target was missed (.1471). Separate analyses were done for each foil lag, but there was no apparent effect at any lag. It might not be wise to draw any far-reaching conclusions from these data, but they do suggest that in the present experiment the "units" in memory were the individual elements of the TBRI. It is quite possible that under typical serial recall conditions a quite different result might be obtained.

The effect of retention interval was powerful. In Figure 14 are the separate retention functions (probability of correct as a function of time) for the different lag conditions. Foil lags 4, 8, and 12 have been combined to reduce visual confusion. For some reason the

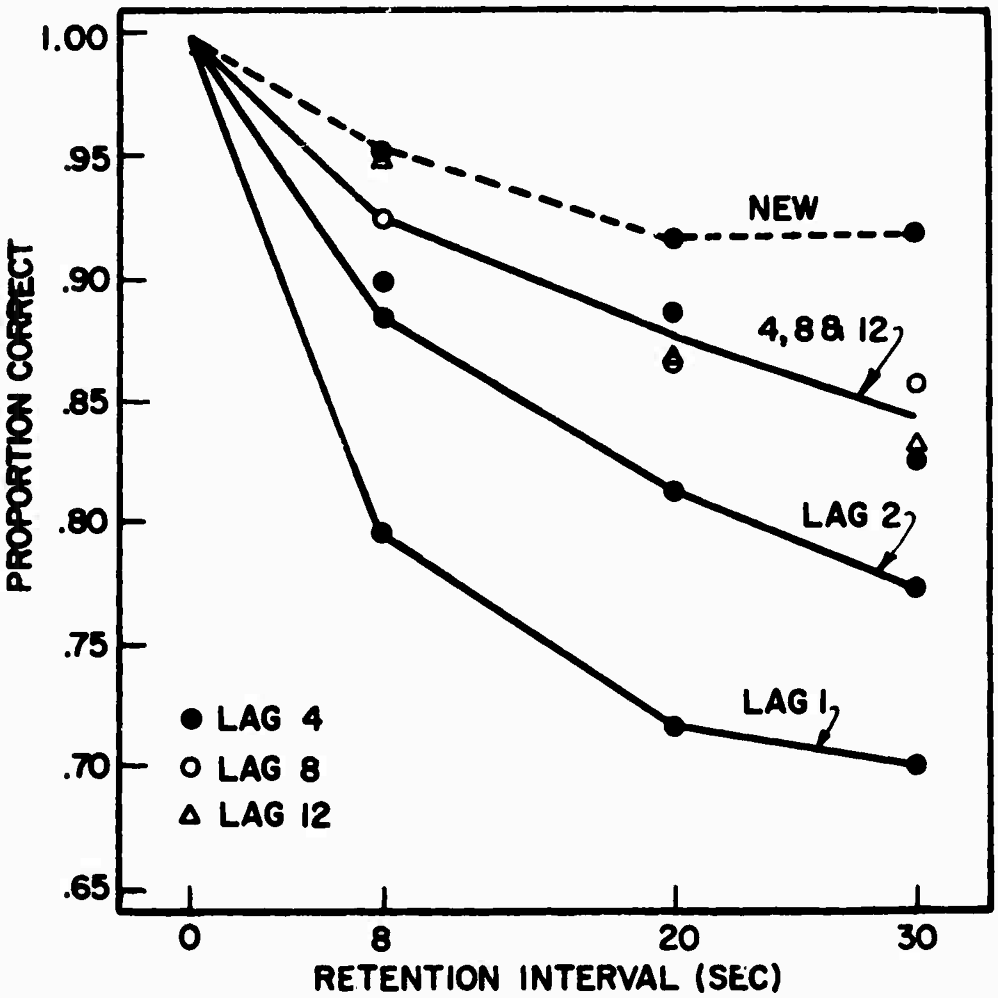

Fig. 14 Retention functions for the different foil lag conditions. The curve labelled 4, 8, & 12 gives the mean value of these three lags. There are 324 observations per data point.

data at long foil lags were instable, despite the fact that each point represents 324 observations (3 observations on each of the 108 Ss.) An analysis of variance was performed with foil lag (6), retention interval (3) and Ss (108) as factors. The effect of foil lag is significant ($\underline{F}_{5,535} = 33.1908$, $\underline{p} < .001$) as is the retention interval effect ($\underline{F}_{2,214} = 28.8001$, $\underline{p} < .001$). However, the retention interval by foil lag interaction is not significant ($\underline{F}_{10,1070} = .9384$). The absence of a significant interaction reflects the similar slopes of the retention functions of all of the intra-experimental foils. From these data it appears that the retention function reaches asymptote at 20 sec when the foil is a new word. It may be the case that the intra-experimental foils would produce retention functions with asymptotes, but if so the asymptotic value is not approached closely until after 20 sec. (Separate ROCs were constructed for each of the foil lag by retention interval conditions, and analyses performed on A_G after applying an arcsin transform. The results were essentially similar to those reported for the raw error data.)

Since the retention interval was a within-S factor, and since other intervals were of constant duration, the time between the presentation of the foil and the TBRI of the current trial varied. In light of the data on temporal spacing (e.g., Loess & Waugh, 1967) it would be expected that this interval would be a determinant of recognition performance. Specifically, foils should be more attractive lures if they are recent. In Figure 15 performance on

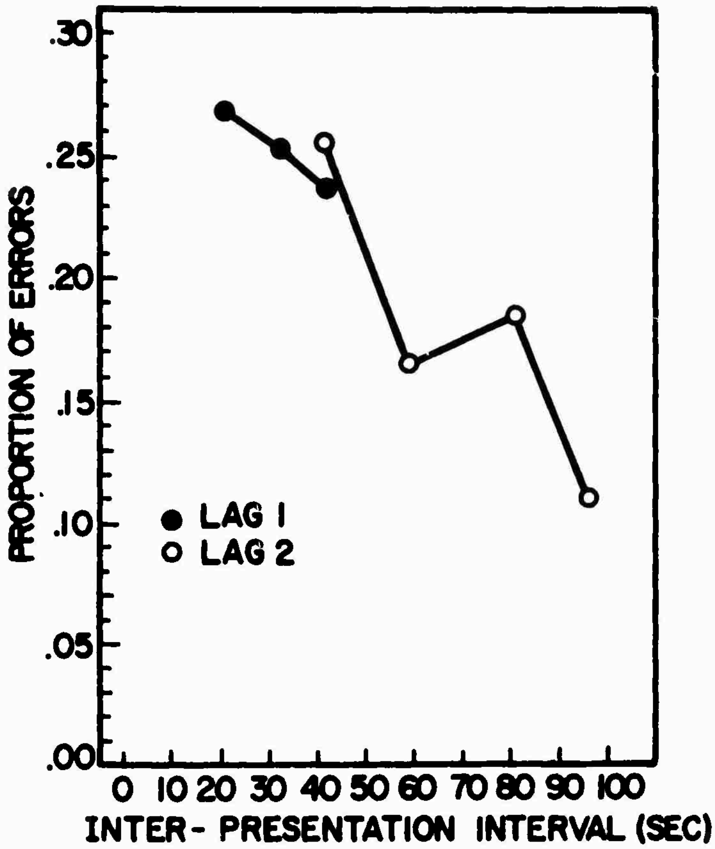

Fig. 15. Error rate as a function of interpresentation interval.

foils of Lags 1 and 2 has been broken down according to the actual time since the presentation of the foil. As can be seen, error rate is strongly influenced by the age of the foil; within either lag condition performance is best when the foil is old. The data for the other lag conditions are considerably more variable. At lags greater than 2 the retention interval of the current trial and the total inter-presentation interval are not orthogonal, and there are a large number of different inter-presentation intervals. Several different analyses were performed, however, and for all lags there was a strong tendency for old foils to be associated with lower error rates.

As was seen in Chapter I, intrusions tend to be serial position specific; most intrusions occur in the same within-item serial position as they occupied during their original presentation (see Table 1). An analogous effect on recognition performance was not obtained in the present experiment. In Table 2 are the matrices which show error rate as a function of foil within-item serial position. There was no tendency for foils to be accepted more readily when their within-item serial position matched the within-item serial position of the target. This could mean that serial position information is not included in the original encoding of the TBRI. It would not be unreasonable to find no serial position coding here, since in this situation Ss are never required to specify order information. However, these same data are equally consistent with the notion that there is serial position encoding, but that such information plays no part in the decision making which occurs at the time of test. It is not clear that the probability of a correct choice would in any way be affected if S knew that both alternatives had the same or different serial positions. On purely logical grounds, it is not obvious that such information could either hinder or help recognition performance. (It would be possible to imagine some fairly complicated inferences which could use this information as an aid to recognition. For instance, an S might realize that one alternative (say "PATH") had been the first word in an item and that the other ("NOON") had been the third.

TABLE 2

ERROR FREQUENCY AS A FUNCTION OF WITHIN-ITEM SERIAL POSITION OF THE TARGET AND OF THE FOIL

Foil Serial Position	Foil Lag 1 TBRI Serial Position 1	2	3	Total	Foil Lag 2 TBRI Serial Position 1	2	3	Total	Foil Lag 4 TBRI Serial Position 1	2	3	Total
1	21	15	40	76	14	22	25	61	10	12	14	36
2	35	26	49	110	14	17	22	53	11	18	19	48
3	13	21	33	67	21	17	20	58	11	18	11	40
TOTAL	69	62	122	253	49	56	67	172	42	48	44	124

Foil Serial Position	Foil Lag 8 TBRI Serial Position 1	2	3	Total	Foil Lag 12 TBRI Serial Position 1	2	3	Total
1	20	12	8	40	11	10	15	36
2	7	14	20	41	14	13	14	41
3	7	12	15	34	13	8	10	31
TOTAL	34	38	43	115	38	31	39	108

If he then remembered that the item in the third serial position of the TBRI was "LAKE," he might select "PATH" as being the only alternative consistent with this information.)

In pilot studies that have not been reported here, there were no signs of changes in performance as a function of the number of prior trials, excepting a slight practice effect. The situation in the present experiment is somewhat more complicated. Performance on recognition tests which used new words as foils is plotted, as a function of trials, in Figure 16. As can be seen, performance on the first trial of the experiment was quite good: there was only 1 error out of a possible 108 (106 Ss were correct with highest confidence, one S was correct with next to highest confidence, and one S was wrong with highest confidence.) This was followed by a rapid build-up in the false alarm rate which gave way to a gradual improvement that continued over the rest of the experiment. This result is surprising. It seems likely that the incorrect acceptance of new foils is produced by processes similar to those responsible for false alarms (calling a "new" word "old") in continuous recognition memory (e.g., Shepard & Teghtsoonian, 1961; Melton, Sameroff, & Schubot, 1967). However in continuous recognition memory experiments there is typically a slow steady build-up of false alarms which seems to continue over several hundred trials. In the context of these experiments, the build-up in false alarm rate observed here is almost unbelievably precipitous. However, in typical continuous

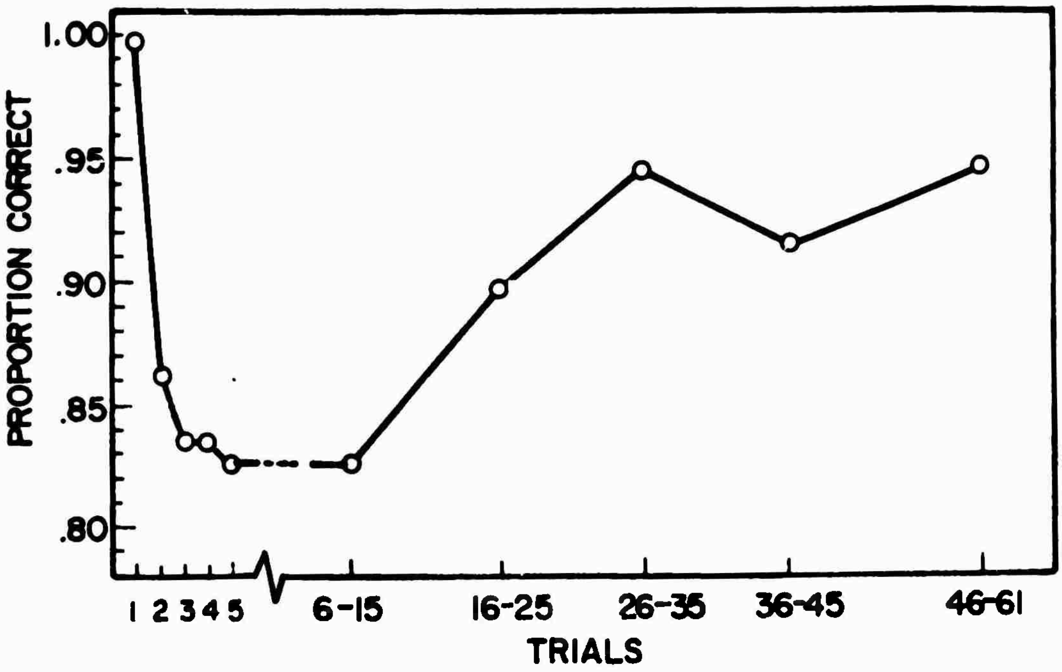

Fig. 16. Probability of correct recognition, as a function of trials, for tests using "new" words as foils.

recognition memory experiments the test is Yes/No rather than 2-AFC. The possibility that this is a crucial difference receives some support from an experiment by Shepard and Chang (1963). They used number triple TBRIs with a 2-AFC test in an otherwise standard continuous recognition memory experiment. While there are not sufficient data to look at early performance on a trial by trial basis, it is the case that the probability of a correct response decreases from the first block of 10 trials to the second, and thereafter remains constant.

The experimental design does not permit an examination of early trial performance on tests using intra-experimental foils. The only proactive effects were practice effects comparable in magnitude to the practice effect shown in Figure 16. It would be very interesting if the intra-experimental foils were also more attractive lures when there were larger numbers of prior trials.

CHAPTER IV

DISCUSSION

There is a striking similarity between the present experiment and conventional psychophysical techniques. In both cases Ss are presented with pairs of stimuli and asked to select the alternative which most closely resembles a standard. In this case the standard is defined not by its physical attributes but rather as the most recent event in an ongoing sequence of events. If data like those displayed in Figure 11 were obtained by systematically varying the physical similarity of the foil and the target it would be acceptable to most to say that errors were caused by "confusions" between the target and the foil. If this logic is extended to the present case, it would mean that the traces of recently presented prior items are confusable with the trace of the TBRI. If this were granted, it would be reasonable to conclude that trace confusability is a primary cause of forgetting in short-term memory.

Control Processes and Structural Features

Unfortunately, this conclusion is not an immediate consequence of the data. There are at least two objections which might be raised. First, it could be argued that the foil-lag effect does not reflect a fundamental characteristic of the memory system but is rather tapping some fairly "high level" response strategies. Second, it is possible that the presumption of an intimate relationship between the recognition test used here and typical recall procedures

is inappropriate. Both of these objections are general enough to preclude a rigorous, straightforward rebuttal. However, it is possible to marshall arguments which cast serious doubt on their plausibility.

Determining when performance of a task is dependent upon a fundamental structural characteristic of memory instead of reflecting a strategy is always difficult. Given current conceptualizations of "structural features" and "control processes" it is not possible to arrive at any formally meaningful distinction. That is, it is never possible to look at the mathematical statement of some particular model and say that some expression or expressions are derived from assumptions about control processes and some others from structural assumptions. Whenever a particular model is modified by changing the assumptions about the strategies employed by Ss it becomes another model. There are no simple rules which determine whether a transformation of a model represents an alteration of control processes assumptions or of assumptions about structural features. Given this ambiguity, it is doubtful that it would ever be possible to demonstrate that some experimental effect was of necessity reflecting a structural characteristic of memory rather than a strategy difference.

In the current case the most that can be said is that it does not appear that any current memory model can handle these data unless it makes provision for confusions about trace identity. These confusions may arise as a result of structural characteristics of the store, or they may be induced by strategies. Instead of

attempting to give a very general argument in support of this conclusion it will suffice to talk about two classes of models that appear to almost exhaust the current models not incorporating confusions between traces. The first class (all-or-none forgetting models which assume no possibility of mistaken identity as a source of errors) is exemplified by the dual-trace model of Chapter I. All models of this sort produce an expression for the probability of correct recognition of the form:

$$p(\underline{C}) = \underline{p}^* + \underline{a}(1-\underline{p}^*),$$

where $\underline{p}^*$ is the probability that the trace of the TBRI is still in a remember state, and $\underline{a}$ is the probability of a correct guess given that the TBRI has been forgotten. The argument demonstrating the inability of the dual-trace model to mimic the predictions of the competition model (Chapter II) was actually general enough to be applied to all models of this class.

The second general class also assumes that the trace representing the TBRI is known to the retrieval system, but allows traces to be partially forgotten. This class of models predicts that recognition performance will be:

$$p(\underline{C}) = \underline{p}^*\underline{b} + \underline{a}(1-\underline{p}^*),$$

where p^* is now the probability that the retrieval system decides to base a response on the trace of the TBRI, which produces a correct response with probability $\underline{b}$, and $\underline{a}$ is the probability of a correct response given that the system opts to guess. In this case, it is

again not possible to predict the foil-lag effect by allowing a to vary. It can also be shown that prediction of the foil-lag effect is not possible if b is allowed to vary. Actually, changes in b with foil lag are quite reasonable, but any model which permits b to vary with foil lag acts just like one which is "truly" confused about which trace in store corresponds to the TBRI. This follows since it can be demonstrated that if b varies with foil lag, then the trace must contain fallible information indicative of the time of storage. Given this, and assuming that the same model is to be applied to the recall situation, there will be confusions in recall between the trace of the TBRI and the trace of prior items.

The most reasonable conclusion from these arguments is that any memory model of the kinds currently available will have to explain the data of the current experiment by appealing to confusions between traces. This does not rule out the possibility that some future memory model might be able to predict these data by assuming no possibility of mistaking some prior item for the TBRI, but postulating relatively complicated response strategies. However, there seems to be no reason to prefer that kind of model over models that postulate trace confusability. So, until there is good reason to doubt the plausibility of assuming trace confusability, it seems most reasonable to concentrate theoretical explorations on "competition" models.

A second possible objection to concluding from the current experiment that memory failure is at least in part attributable to

confusions among traces is that the recognition test used here may not be comparable to the typical recall procedure. Arguments about the generalizability of a result can become exceedingly long-winded, but no such elaborate defense will be provided. It is worth noting, however, that given the apparent confusability of traces in the present experiment, it is only reasonable to assume that intrusions are the result of a similar confusion in recall attempts. It would be unnecessarily complicated to accept trace confusability for the recognition test and then explain apparent confusability in recall by appealing to an alternative mechanism. It might be noted that accepting this conclusion confers an additional benefit. Proactive interference is one of the most powerful determinants of retention in the B-P situation. If the notion that recently created traces are confusable with the trace of the TBRI is accepted, then PI is seen to be a natural consequence of the structure of memory.

In summary, it seems reasonable to conclude that, in the B-P situation, confusions about which trace represents the TBRI are one of the major causes of forgetting. This conclusion in turn suggests that while a model like the dual-trace model may, under some circumstances, provide a reasonable description of the data, it fails to reflect the processes responsible for forgetting. While "steady-state" models like the dual-trace model may serve a useful purpose, it is likely that a sufficient account of memorial capacities in situations similar to the B-P will have to be constructed with a very different framework. The major import of this conclusion is

not that it denies the dual-trace model. While the dual-trace model does a respectable job of handling the data from B-P experiments, it is not remarkable to find yet another promising model incorporating assumptions that turn out to be quite unreasonable. Of much more interest, most of the landmark theoretic papers in the short-term memory literature take no account of PI and propose models that make no provisions for trace confusability (e.g., Atkinson & Shiffrin, 1968; Bower, 1967; Brown, 1959; Norman & Rumelhart, 1970). In a sense, the conclusion that trace confusability mechanisms are essential is in fundamental disagreement with the assumptions that underly the most influential of the current quantitative models of forgetting. The remainder of this Chapter will be devoted to considering various models that might be used to capture the notion of trace confusability.

Models of Trace Confusability

Classical Strength Theory

If the analogy between the present experiment and psychophysical procedures is accepted, then it would seem reasonable to turn to psychophysical models for an explanation of the data. One possibility is to imagine that at the time of test the S examines the traces corresponding to the two presented alternatives and attempts to determine which is most recent. Following the theory of signal detectability (TSD), it can be supposed that recency is represented by some unidimensional attribute subject to random variation. At the time of test it is assumed that the momentary values of the recency attributes of the traces corresponding to the presented

alternatives are compared, and the alternative whose trace appears most recent is selected as the response. The probability of a correct response is just $p(\underline{T}_t > \underline{T}_f) = p(\underline{T}_t - \underline{T}_f > 0)$, where $\underline{T}_t$ and $\underline{T}_f$ are, respectively, the momentary perceived recency of the trace of the target and of the foil. If it is assumed that $\underline{T}_t$ and $\underline{T}_f$ are normally distributed random variables, then their difference, $\underline{T}_d = \underline{T}_t - \underline{T}_f$, will also be normally distributed; consequently, the probability of a correct response is just the integral from 0 to infinity of the normally distributed random variable, $\underline{T}_d$.

To explain changes in the probability of a correct response as a function of foil lag it will be necessary to specify an additional function that describes the way in which apparent recency of traces changes with time since presentation. It would then be possible to apply this model to either the data of the current experiment, or to data from experiments using recall tests. (While extending this model to recall, or N-alternative recognition memory, is possible, the decision rule is more difficult to model. It is almost always impossible to arrive at an analytic solution, although there are some accurate and practicable numberical estimation procedures.)

Interestingly, the preceding model uses a decision rule identical to that commonly used in interference theory analysis (e.g., Spence, 1936). In the competition model derived from interference theory in Chapter I, it was assumed that at the time of presentation

the trace of the TBRI is associated with the current context. In order to induce some reasonable amount of variability in performance, it is typical to assume that associative strengths undergo random oscillation. This notion is usually formalized by assuming that the strengths of associations are random variables with normal density functions. The decision rule is to accept as the preferred alternative the one whose trace has the strongest momentary association to the context. Clearly, the probability of the target being accepted in a 2-AFC test is $p(\underline{S}_t - \underline{S}_f > 0)$, where $\underline{S}_t$ is the strength of the association of the context with the target, and $\underline{S}_f$ is the strength of the corresponding association with the foil. This probability is again the integral from 0 to infinity of a random variable, $\underline{S}_d = \underline{S}_t - \underline{S}_f$, with a Gaussian distribution.

Either of the preceding formulations (which seem to differ only in verbal description and connotations) might provide a reasonable description of performance in the B-P task. All that is lacking for a complete model is a specification of the function relating apparent recency (or trace strength) to the actual age of a trace (or association). The data in the present experiment provide some constraints on the nature of this function. Specifically, the decision rule postulated by the two models is such that it is possible to solve for the values of the trace strength at each retention interval. Any proposed function relating trace strength to time in store must be consistent with these estimated values. (It need not predict the exact values estimated from the data, but

the set of values it generates must be related by a linear transform to the estimated values.) For the decision rules used here, it is possible to solve for the expected value of the recency distribution of the traces using the familiar Thurstone Case V analysis (e.g., Torgerson, 1963). (The important assumptions of this analysis are (a) that the distributions of trace strengths are independent, and (b) that the variance of the distributions do not covary with their mean values.) This analysis was performed and the results are in Table 3. Also in this table may be found the results of a similar scaling procedure which is derived from Luce's choice theory (e.g., Luce, 1962). This method assumes that all alternatives have a certain "valence" or attractiveness, $\underline{V}_{\underline{i}}$, and that the probability of selecting alternative $\underline{a}$ over alternative $\underline{b}$ in a 2-AFC test will be:

$$P(\underline{a},\underline{b}) = \frac{\underline{V}_{\underline{a}}}{\underline{V}_{\underline{a}} + \underline{V}_{\underline{b}}} .$$

In general, this decision rule will produce results quite similar to those obtained from the interference theory decision rule (Luce & Galanter, 1963), but it is considerably easier to handle when it is necessary to deal with multiple-alternative recognition tests (or, equivalently, with recall). The general form for selecting one alternative from some set $\underline{R}$ is just

$$P(\underline{a}:\underline{R}) = \frac{\underline{V}_{\underline{a}}}{\sum_i \underline{V}_{\underline{i}}}$$

where the sum ranges over all the alternatives in $\underline{R}$.

For both of these scaling procedures it was assumed that the strength (or apparent recency) of a new item would not change over the retention interval. Consequently, its strength was set equal to one, and all of the other strengths estimated under this assumption. As can be seen by an inspection of Table 3, the strength of the

TABLE 3

ESTIMATED TRACE STRENGTHS

	Strength Theory			Luce's Theory		
	Retention Interval			Retention Interval		
Alternative	8	20	30	8	20	30
TBRI	1.69	1.39	1.40	14.38	11.05	11.50
Lag 1	0.87	0.78	0.88	3.75	4.13	4.78
Lag 2	0.50	0.51	0.65	1.91	2.56	3.26
Lag 4	0.42	0.14	0.65	1.63	1.68	2.36
Lag 8	0.27	0.28	0.33	1.20	1.94	1.83
Lag 12	0.00	0.28	0.30	1.00	1.65	2.06
New	0.00	0.00	0.00	1.00	1.00	1.00

TBRI declines over the retention interval, but the strengths of the other alternatives tend to increase. The implication of this result is clear. If the decision rule used by interference theory (or TSD) is adopted, then it will be necessary to postulate something like "spontaneous recovery."

While this is not necessarily an embarrassment to interference theory (in fact, Keppel & Underwood, 1962, originally proposed that the traces of past items undergo experimental extinction when the TBRI is presented and then spontaneously recover over the remainder of the retention interval), it does weaken the temporal judgment model. It is difficult to see why apparent recency would be anything other than monotonic with time.

The foregoing analysis cannot, of course, be accepted as strong support for the notion of spontaneous recovery. The estimated strengths in Table 3 entail spontaneous recovery only if the decision rule postulated to arrive at the scale values is accepted. A different decision rule which would produce monotonically decreasing trace strengths could have been assumed. However, the inference of spontaneous recovery could be strengthened somewhat if it could be shown that the decision rule used here is particularly appropriate to these choice data. In the preceding analysis there was no attempt to predict the choice data, in fact there was just sufficient data to produce an estimate of the strength of each trace. However, if an experiment were conducted in which choices were made between all possible pairs of alternatives (including cases where both alternatives were wrong), it would then be possible to estimate trace strengths from part of the data, and use these strengths to predict the remainder of the choices. If either the interference theory decision rule, or Luce's decision rule, gave good predictions of the choice probabilities, and if the estimated

strengths of the traces recovered over the retention interval, then it would be harder to avoid the assumption of spontaneous recovery. The only way to avoid that assumption with a "strength" model would be to display another decision rule which was equally competent at describing the pattern of choice probabilities but which did not require trace strengths to increase over the retention interval.

As an aside, it should be noted that there are a large number of current models that assume decision rules like Luce's or the one proposed by interference theory. The adequacy of the decision rule is in almost all cases testable by applying some scaling procedure, but no one appears to have done so. This is a regrettable omission since decision rules are crucial components of models and are hardly ever testable except by attempting to predict sets of choice probabilities.

While models like the preceding one certainly warrant further consideration, the data from the current experiment are not particularly appropriate for evaluating them. The decision rules in models of that sort will almost inevitable by fairly complicated, and the current data do not provide any way to compare the relative merits of different rules. Consequently, it seems more appropriate to survey the field of possible models in an attempt to locate one that appears simple but that also seems to capture the essential features of "trace confusability." The difficulty with this scheme is that it is necessary to decide which features of trace confusability are indispensible and which represent more or

less accidental products of relatively unimportant assumptions. Decisions about the relative centrality of different assumptions rely upon an overall conceptualization of the memory system. Since little effort has been devoted to constructing models that use trace confusability as the prime cause of forgetting, it is perhaps not surprising to find that no commonly known scheme will do the trick. Consequently, the first step is an attempt to generalize currently popular types of models and relate these to a speculative structure which includes all of them as special cases. Once the overall structure has been sketched out, discussion will turn to a specific model which seems to represent the simplest possible specialization.

An Overview

The first point to be made is that the competition and trace-degradation models of Chapter I share an interesting structural similarity. Conrad's trace-degradation model is actually somewhat more involved than would be thought from either his statement of it (e.g., Conrad, 1960, 1967) or from the discussion of Chapter I. It is assumed that the retrieval system examines the fading trace of the TBRI and constructs a response on the basis of the information available there. It is actually necessary to postulate a rather complicated trace reconstruction process if the model is to mimic human behavior. Most notably, the retrieval system does not just generate a response consistent with the information in the trace, but rather produces an acceptable English word. It is conceivable that an S might report "TEST" when the TBRI has been "BEST."

However, he would never replace "BEST" with "BESB" or "BESP," even though they also represent simple one phoneme transformations of the original stimulus. It seems inevitable that any system like Conrad's will have to supply the retrieval system with a "dictionary" that specifies possible responses. (Conrad may well be aware of this requirement. The principle data to which his theory has been applied are the recall of strings of consonants, where the vocabulary is small and known. Conrad's model can be described as one which selects possible responses from the alternatives in the dictionary by looking for entries which have the characteristics specified by the trace. (Conrad originally imagined that the trace was largely devoted to phonemic information, but it is possible to assume that it contains a more inclusive description of the TBRI without doing violence to the basic structure.)

The competition model can be described in similar terms. Here it is assumed that dictionary entries are marked by associations with the context, and that the entry used as a response is selected by choosing the one most strongly associated with the experimental context. If this overall structure is accepted, it is obvious that the competition analysis derived from interference theory and Conrad's trace-degradation model are very similar notions. The most important difference between them is in the criteria they assume to be used in selecting a response. The trace-degradation model assumes that the trace contains information describing the presented

material while competition theory assumes that the stored information is designed to specify the conditions under which the material was presented. It would clearly be possible to devise a "super-model" which included both as special cases. It would provide mechanisms to screen responses on the basis of their associations with the context (i.e., "recency" information) and on how well they fit with some dictionary description of the TBRI.

Both the competition and the trace-degradation models assume that the "trace" of the TBRI represents not a description of the response _per se_, but rather a statement of how to find the correct response in some permanent memory. For both models it seems to be necessary to distinguish information which is a permanent part of the organism's repertoire from information about particular past events. Both models also assume that memory failure is the result of a gradual loss of information about the dictionary entry originally indicated by the TBRI. However, it is possible to imagine a model which assumes that traces are not subject to information loss, but instead accounts for memory failures by postulating variability from different sources. Two such models will be described.

It is possible to accept the basic structure of the competition model while assuming that memory loss is attributable not to changes in associative strength, but rather to alterations in the context to which responses are associated. It was assumed that at the time of presentation the _current_ context is sampled, and the response most strongly associated to this context is selected as correct.

In earlier discussions of this model it was assumed that the context is relatively invariant and that memory failures are produced by changes in the strengths of the associative connections between responses and the context. It would be equally appropriate to assume that there is no loss of associative strength, but rather a shifting of the context. If at the time of recall the context were more similar to that present on some prior trial than it was to the context of the current trial, the retrieval system might erroneously decide upon some prior TBRI as the most appropriate response. In a model of this type, it would be possible to predict memory failures even though no information was being lost from store. All forgetting would be because the memory system was coupled to a "noisy" environment.

An alternative structure could share with the preceding model the assumption that there was no loss from store, but rely upon changes in "encoding" to produce errors (cf., Martin, 1968, 1971). This type of model is most easily described for the recognition situation. In its simplest form it would say that a "miss" occurs because the target is not encoded in the same way as it was on its original presentation. A "false alarm" would be produced when an incorrect alternative was encoded in a way which made it appear similar to the trace of the correct response. A crucial point for models of this sort is the aspect of encoding which is subject to variability. Under normal conditions, it is safe to assume that an identical word will produce the same encoding, if by encoding

all that is meant is arriving at an understanding of the natural language usage of the word. Obviously, variability in encoding will have to occur after the word's natural language meaning has been identified. To be a bit philosophical, a distinction is necessary between recognizing the "public" meaning of a word and realizing that a particular verbal event corresponds to an earlier "private" experience. Quite clearly, a precise statement of what is meant by encoding variability will have to wait upon a more detailed analysis of the transformations made upon internal representations of events.

At first glance, it would appear that the preceding conceptualizations of memory have little in common. While it is certainly true that there are some substantive differences in what has been assumed about memorial processes, there also appears to be considerable underlying agreement about the basic structure of memory. Specifically, all of these models suggest that "memory" really refers to two very different sorts of information conservation. On the one hand, to remember a word may mean to have available information corresponding to a dictionary definition. On the other hand, it may mean to remember that this particular verbal item occurred in some particular context at some particular time. There are many ways in which this distinction might be expressed. For the purposes of this paper, however, the terms "object code" and "event code" will do nicely. By an object code will be meant a statement of the public meaning of a word. For instance, the

object code for "CAT" would include a listing of the commonly agreed to characteristics of cats (e.g., small, furry, self-possessed carnivore) and probably some statement of how the word may be used in a sentence. It is assumed that whenever the word "CAT" is presented, essentially the same object code will be generated. (There may well be some differences in the particular features of a word that are highlighted, but it will be assumed that, for experiments of the type considered here, this is not an important determinant of retention.) An event code is a record of the occurrence of some particular word on some particular occasion.

The most important differences among the models discussed here are in the assumptions they make about what an event code is and how it is used in reconstructing past experiences. It may be useful to summarize the assumptions of the various models and make some specific comments about their inter-relations. For the original version of Conrad's trace-degradation model it would appear that an event code contains a fading representation of the object code of the presented material. It is presumed that there will never be any difficulties in arriving at the appropriate event code. All forgetting will be the result of the event code having become so illegible that it no longer corresponds to a single object code.

With the competition model it is assumed that the event code will always suffice to lead back to the appropriate object code. Forgetting occurs because of the retrieval system's inability to locate the appropriate event code. It is presumed that event codes

have associated with them some strength value and that the retrieval system's best guess about which event code is the correct one is always that it is the event code with the largest associated strength. The contextual variation version of the competition model is very similar. However, it assumes that event codes contain something like a statement of the context under which the event occurred. At the time of recall, the current context is sampled, and the event code with the most similar contextual marking is selected. It should be noticed that the contextual variation model and the "standard" interference theory version of the competition model are formally very similar. There is no obvious difference between specifying the way in which the context changes versus specifying the manner in which associative strengths change. Both of these notions can be expressed in terms of some function of time which specifies the likelihood that a particular trace will be accepted as the trace of the TBRI. However, it is probably the case that the range of permissable functions is somewhat larger for the context model. There are certain restrictions which would probably be placed on the way in which associative strengths change (e.g., as continuous functions of time) that might sensibly be relaxed if one were talking about context changes.

The encoding variability model is also quite similar to these other models. For recognition tests, it assumes that forgetting is the result of encoding the recognition alternatives in a way that is different from the manner in which they were encoded at

original presentation. This would best be interpreted as asserting that the event codes generated at the time of test are different from the event codes used to store the material originally. (It is again assumed that there is no important variability in the generation of object codes.) Then it is reasonable to assume that if the appropriate event code is located, recognition will be prefect. As might be expected, the formal structure of the model is very similar to that of the competition model. In this case, the basic determinant of memory will be the probability that the system will return the same event code at the time of test as it did at the time of presentation. This again would be specified as some function of time.

It should be noted that the contextual-variability model is a special case of the encoding-variability model. The contextual-variability model asserts that the context is part of the event code; failures of recognition (or recall) occur because changes in the context produce changes in the event codes generated by the system. In other words, there are changes in encoding induced by alteration of the context. The general notion of encoding-variability suggests that changes in the "context markings" of an event code are only one of a large number of possible sources of encoding variability. In its most general form encoding-variability notions permit contextual changes to change not only event codes but also object codes.

The remainder of this Chapter is devoted to a quantitative model for recognition memory as exhibited in the experiment reported here. A model that appears to be the simplest possible specialization of the general structure discussed above is introduced, and the inadequacies of the model are examined with the intent of locating potentially fruitful ways of shaping the model into a more comprehensive structure. The formal statement of the model can be interpreted in many ways: It was designed to be describable as an exemplar of associative interference theory, or as a contextual-variability model, or as one of the more general encoding variability schemes. To avoid foreclosing any of these interpretations a more or less neutral terminology is adopted. Instead of speaking of associative strengths (or commonality of context) the term "recency information" is used. The term is meant to denote whatever information is used to determine when and where the trace (event code) was created.

A Simple Two-State Model

The simplest temporal discrimination among traces would obtain if event codes could be partitioned into only two classes, "recent" and "old." If this were the case, the decision rule used in generating a response could be very simple. For example, it would be logical to select the alternative whose event code appeared more recent, or to guess randomly if both event codes were of the same apparent age (both "recent" or both "old"). This very straightforward assumption yields a simple model. To state the model, let $p(C_{i,t})$

be the probability of a correct response when the foil is of lag $\underline{i}$ and the retention interval is $\underline{t}$ sec, let $\underline{P}_{0,t}$ be the probability that the TBRI is still perceived as recent after a $\underline{t}$-sec retention interval, and let $\underline{P}_{i,t}$ be the corresponding probability for a foil of lag $\underline{i}$. In terms of this notion the probability of a correct response is

$$p(\underline{C}_{i,t}) = \underline{P}_{0,t}(1-\underline{P}_{i,t}) + 1/2\underline{P}_{0,t}\underline{P}_{i,t} + 1/2(1-\underline{P}_{0,t})(1-\underline{P}_{i,t}),$$

which reduces to:

$$p(\underline{C}_{i,t}) = 1/2 + 1/2(\underline{P}_{0,t} - \underline{P}_{i,t}). \qquad (1)$$

Since there are only two recency states, it is reasonable to assume a simple geometric loss process with Markov properties. Thus the transitions from the recent (R) to the old (Ø) state might be goverened by a Markov process with transition matrix $\overline{T}$ as below

$$\overline{T} = \begin{array}{c c} & \begin{array}{cc} R & \emptyset \end{array} \\ \begin{array}{c} R \\ \emptyset \end{array} & \begin{bmatrix} 1-f & f \\ 0 & 1 \end{bmatrix} \end{array} \qquad (2)$$

This matrix would lead to an expression for $\underline{P}_{0,t}$ (the probability that the TBRI is still recent) of the form:

$$\underline{P}_{0,t} = (1-\underline{f})^{t}.$$

It would be simplest if the probability of a foil still being perceived as recent ($\underline{P}_{i,t}$) could be defined analogously, say as $\underline{P}_{i,t} = (1-f)^{t'}$, where $\underline{t}'$ is the total time elapsing between the original presentation of an item and its occurrence as a foil on

a recognition test. However, as might be expected, this is not a workable assumption. The problem is that the time between the presentation of an item and the recognition test of that trial is filled with a relatively homogenous activity, the interpolated task. However, between the presentation of an item and its re-presentation as a foil, there are not only retention intervals, but also study periods, tests, and rest intervals. There are many ways in which the non-homogenity of the activity might be handled within the model, but instead of trying to develop an elaborate system, the apparent age of a trace can be assumed to decrease a constant proportion for every trial that intervenes between its presentation and the time it appears as a foil. This constant, $\underline{r}$, can be set equal to $(1-\underline{f})^{w}$, and $\underline{w}$ estimated directly for the data. This leads to a relatively simple expression for $\underline{P}_{i,t}$:

$$\underline{P}_{i,t} = (1-\underline{f})^{t+wi}. \tag{3}$$

The complete formal statement of the model can be arrived at by substituting the values for $\underline{P}_{0,t}$ (Equation 2) and $\underline{P}_{i,t}$ (Equation 3) into the expression for $p(\underline{C}_{i,t})$ (Equation 1). The resulting expression can be reduced to Equation 4:

$$p(\underline{C}_{i,t}) = 1/2 + 1/2(1-\underline{f})^{t}(1-(1-\underline{f})^{wi}) \tag{4}$$

In Table 4 are the predicted values generated by this model and the observed recognition probabilities for all intra-experimental foils at all retention intervals. As can be seen, the model provides a

TABLE 4

OBSERVED AND PREDICTED VALUES FOR INTRA-EXPERIMENTAL FOILS AT ALL RETENTION INTERVALS (F = .008, w = 70)

Retention Interval	Foil Lag 1 O	1 P	2 O	2 P	4 O	4 P	8 O	8 P	12 O	12 P
8 Sec.	.027	.211	.117	.110	.102	.063	.077	.057	.046	.057
20 Sec.	.272	.259	.188	.175	.105	.136	.133	.131	.130	.130
30 Sec.	.302	.293	.228	.221	.176	.187	.142	.182	.157	.182
χ^2	.3415		.5608		10.0630		5.2180		1.743	

the model provides a credible description of the data. The overall χ^2 is 17.9263, which, with 13 degrees of freedom (15 less the two estimated parameters), yields a p between .15 and .20. It should be noticed that most of the discrepancy between the obtained and expected values of χ^2 is attributable to a single point (performance after an 8-sec retention interval with a foil of Lag 4). As can be seen in Table 4 this point appears to be seriously out of place. Given this anomalous data point it is not likely that any sensible model would yield an appreciably better fit to these data.

While the model does do a reasonable job of predicting performance for intra-experimental foils, there are several places where it encounters difficulties. The first is in attempting to predict the way in which Ss use confidence ratings. As can be seen from the Receiver Operating Characteristics plotted in Figure 12, Ss

seem to be using the confidence ratings in a sensible manner. When the data are pooled over Ss, it is always the case that an increase in confidence is associated with an increased probability of a correct response. As the proponents of the Theory of Signal Detectability have been pointing out for the last decade, there is no obvious way to account for Ss using more confidence ratings than there are detection states, if confidence ratings really reflect the detection state triggered by the presentation of a stimulus. This implication is certainly the case for the current model. In Table 5 are the observed and predicted confidence ratings for each foil lag condition and each retention interval. The predictions were made by assuming that the S would be certain (use a rating of 4) when one alternative was associated with a recent event code and the other with an old event code. Otherwise he would hapharzardly select one of the other confidence ratings (3, 2, or 1).

Quite clearly, the predictions in Table 5 do not constitute a reasonable description of the data. While the trends apparent in the use of high and intermediate confidence are the same for the data and the model, the overall agreement is quite poor. Although it is certainly desirable to be able to account for the confidence rating data, it is not the case that the model need be dismissed because of a failure to do so. Krantz (1969) has shown quite clearly that "threshold models" (i.e., models with limited numbers of detect states) are contradicted by the systematic use of confidence ratings only if it is assumed that there is a deterministic relation between

TABLE 5

PREDICTED USAGE OF CONFIDENCE RATINGS

Foil Lag 1

		Hit 4	Hit 3	Hit 2	Hit 1	False Alarm 1	False Alarm 2	False Alarm 3	False Alarm 4
8 Sec.	Obs.	.596	.111	.065	.022	.034	.022	.056	.096
	Pred.	.614	.059	.059	.059	.059	.059	.059	.035
20 Sec.	Obs.	.466	.127	.077	.056	.052	.052	.074	.096
	Pred.	.549	.064	.064	.064	.064	.064	.064	.067
30 Sec.	Obs.	.407	.142	.071	.077	.068	.080	.062	.093
	Pred.	.495	.071	.071	.071	.071	.071	.071	.080

Foil Lag 2

		Hit 4	Hit 3	Hit 2	Hit 1	False Alarm 1	False Alarm 2	False Alarm 3	False Alarm 4
8 Sec.	Obs.	.682	.114	.056	.031	.025	.028	.028	.037
	Pred.	.791	.033	.033	.033	.033	.033	.033	.012
20 Sec.	Obs.	.534	.133	.062	.083	.056	.049	.034	.049
	Pred.	.673	.051	.051	.051	.051	.051	.051	.023
30 Sec.	Obs.	.503	.111	.105	.052	.031	.062	.068	.062
	Pred.	.587	.064	.064	.064	.064	.064	.064	.028

Foil Lag 4

		Hit 4	Hit 3	Hit 2	Hit 1	False Alarm 1	False Alarm 2	False Alarm 3	False Alarm 4
8 Sec.	Obs.	.713	.099	.040	.046	.025	.028	.028	.022
	Pred.	.875	.021	.021	.021	.021	.021	.021	.005
20 Sec.	Obs.	.599	.154	.093	.043	.040	.031	.019	.022
	Pred.	.731	.044	.044	.044	.044	.044	.044	.003
30 Sec.	Obs.	.559	.133	.086	.052	.059	.040	.034	.037
	Pred.	.630	.061	.061	.061	.061	.061	.061	.003

Foil Lag 8

		Hit 4	Hit 3	Hit 2	Hit 1	False Alarm 1	False Alarm 2	False Alarm 3	False Alarm 4
8 Sec.	Obs.	.738	.099	.043	.040	.022	.025	.015	.019
	Pred.	.886	.019	.019	.019	.019	.019	.019	.000
20 Sec.	Obs.	.577	.154	.062	.071	.046	.040	.028	.022
	Pred.	.739	.043	.043	.043	.043	.043	.043	.000
30 Sec.	Obs.	.540	.151	.080	.071	.043	.037	.046	.031
	Pred.	.635	.061	.061	.061	.061	.061	.061	.000

Foil Lag 12

		Hit 4	Hit 3	Hit 2	Hit 1	False Alarm 1	False Alarm 2	False Alarm 3	False Alarm 4
8 Sec.	Obs.	.738	.117	.052	.049	.009	.019	.003	.012
	Pred.	.886	.019	.019	.019	.019	.019	.019	.000
20 Sec.	Obs.	.580	.136	.099	.056	.025	.031	.046	.028
	Pred.	.739	.043	.043	.043	.043	.043	.043	.000
30 Sec.	Obs.	.543	.111	.120	.071	.068	.040	.031	.015
	Pred.	.635	.061	.061	.061	.061	.061	.061	.000

the detection state and the confidence rating selected by the S. It is not at all unreasonable to expect confidence rating data like those obtained here to be produced by a 2-state detection system coupled with a probabilistic decision maker. The assumption of a probabilistic relation between detection state and confidence rating is quite plausible. For instance, in the current case, it would be reasonable to assume that the confidence rating selected was in part determined by the S's estimate of how thoroughly he had learned the TBRI of that trial.

While it is possible to dismiss objections based on the inability of the model to handle the confidence rating data, there are some more serious problems. First, the current version of the model predicts that recognition after a 0-sec retention interval should be relatively poor and strongly dependent upon the foil lag. This prediction is easily seen if t is set equal to 0 in Equation 4:

$$p(C_{i,0}) = 1/2 + 1/2(1-f)^0[1-(1-f^{wi})]$$
$$= 1 - 1/2(1-f)^{wi}.$$

With the parameters estimated from the current experiment this would lead to the expectation that the 0-sec error rate would be, for foils of lags 1, 2, 4, 8 and 12 respectively: .1736, .0603, .0073, .0001, .000002. On the basis of previous pilot studies (and intuition) an effect like this appears unlikely; a good guess would be that 0-sec retention intervals are associated with near perfect performance for all foil lag conditions.

There are two obvious ways in which the model could be altered to make its predictions about performance on an immediate test more realistic. One possibility is to assume three recency states, say R_1, R_2, and $\emptyset$. It could then be assumed that loss from R_1 was relatively rapid. Consequently, immediately after presentation, the trace of the TBRI would be likely to be in R_1, but the event codes corresponding to previously presented material would have dropped down into states R_2 or $\emptyset$. This scheme also provides reasonable fits to the data from the current experiment. The fits can not be any worse than the predictions in Table 4, since if the probability of dropping out of state R_1 were set equal to 1, the three-state model is identical to the two-state model.

A second way would be to adopt a structure like that used by the dual-trace model of Chapter I. It could be imagined that there is a transient representation (perhaps the object code constructed when the item was originally presented) which is not subject to competition. Recognition would then be correct if either the temporary trace were available, or if the correct event code were located. This notion leads to a model that predicts near perfect performance after a very short retention interval. It also must provide fits to the experiment that are at least as good as those displayed in Table 4, since it assumes the form of Equation 4 if the decay rate for the low level trace is assumed to be very high. Consequently, it appears that there is no unavoidable difficulty in modifying the model to predict very good recognition at 0 sec.

It is possible, however, that neither of the schemes presented here would provide a good quantitative description of forgetting across all retention intervals.

The second objection is the most telling. The model seems unable to handle the data from trials involving new words as foils. The most natural assumption is that a new foil is equivalent to a foil of infinite lag, that is, it is always perceived as an "old" item. Under this assumption, Equation 4 takes the form:

$$p(\underline{C}_{\infty,t}) = 1/2 + 1/2(1-\underline{f})^{t}.$$

Unfortunately, this scheme does not work well. In Table 6 are the predicted and observed error rates when the foil was a new word,

TABLE 6

PREDICTIONS OF THE TWO-STATE MODEL FOR TESTS USING NEW WORDS AS FOILS ($\underline{f}$ = .015, $\underline{w}$ = 70)

	Obs.	Pred.
8 Sec.	.046	.057
20 Sec.	.083	.130
30 Sec.	.080	.187

assuming the same parameters used in Table 4. The predictions are wildly off. The predicted error rate is entirely too high. More importantly, the shape of the retention function predicted by the model appears to be wrong. In Table 6 (or Figure 14) the retention

functions for "new foils" appears to asymptote before 20 sec. The predicted retention functions show no such tendency. They continue down until near-chance performance is reached after about 500 sec ($p(C_{-\infty,500}) = .5003$). While it is not possible to be sure that the data from the current experiment reflect a true asymptote, it is certainly true that the loss of information is much less rapid than would be expected from the model.

Table 6 makes it clear, under the assumptions of the current model, that a foil of infinite lag and a word new to the experiment are not equivalent. A word actually presented at any time in the past appears to be a more potent competitor than is a new word. There are several ways in which such an effect might be incorporated into the model. One way is to assume that the event code for every presented item is marked as "intra-experimental" and that this marking is practically permanent. This assumption corresponds to having in store something like the information responsible for "list differentation." An attractive way to state this assumption is by means of a three-state model in which the states might be labelled, IR, IØ, and E. State IR would be the one usually occupied by recently presented material. State IØ would be interpreted as old items which had been presented, and state E would be reserved for material which had never been presented. If it is assumed that the probability of going from IØ or IR to E was quite small, it would be expected that the error rate when the foil was a new word would still be lower than the rates for foils of long lag. (If the

probability of a transition into E is greater than 0, it would still be the case that new foils and foils of infinitely long lag would be equivalent. However, the limiting value is approached much more slowly than it is for the two-state model.) Several models using this general assumption have been developed, and applied with some success to the data from the current experiment. The most promising version assumes that there is some probability that an event code for an item will never be created. If this happens, the only record of the occurrence of an item is the rapidly fading object code. For this model, performance when the foil is a new word declines very gradually after the first 5 to 10 sec. Error rate does not approach chance level until somewhere in the region of 10^6 sec.

However, there is still another difficulty associated with tests using new words as foils. As can be seen in Figure 16, the probability of an error is strongly dependent upon the number of prior trials. Given the mechanisms proposed above it is not obvious why "list differentiation" should become progressively worse over the course of the experiment. An alternative approach can be seen if the assumptions underlying the model are rephrased slightly. It can be supposed that at the time of test there are a number of event codes that appear to represent recent experiences. The retrieval system goes through all of these recent event codes and attempts to match them to the recognition alternatives. The alternative that corresponds to some recent event code is selected

as the correct one. If both or neither of the alternatives match a recent code a random guess is made. If it assumed that matching alternatives to event codes is always performed without error, the decision rule of Equation 1 is generated. However, it is not unreasonable to suppose that the retrieval system could be wrong in deciding which alternative corresponds to which event code. The matching process could be expected to produce both false alarms (incorrectly accepting a recent event code as a match for a "new" foil) and misses (failing to accept the event code of the TBRI as a match for the target alternatives).[3] It is reasonable to expect a model of this sort to predict an increase in error rate with increasing numbers of prior trials. The probability of a false alarm would be expected to increase as the number of event codes marked as recent increased. The model predicts that the number of recent event codes will increase with increasing numbers of prior trials until an equilibrium is reached. The probability that different numbers of events appear to be recent, and the expected number of event codes judged as recent, are shown in Table 7 as a function of trials. It is encouraging that the expected number of recent event codes approaches its asymptote at roughly the same time as the false alarm rate in Figure 16 is approaching its maximum value.

---

[3] To facilitate discussion only tests with new alternatives are considered here. It would be possible, and desirable, to extend these arguments to cover the cases where both recognition alternatives had been presented as TBRIs.

TABLE 7

DISTIRBUTION OF NUMBER OF TRACES IN RECENT STATE (r) AS A FUNCTION OF TRIALS

Trial	p(r=0)	p(r=1)	p(r=2)	p(r=3)	p(r=4)	p(r=5)	Expected Value
1	1.0000						0
2	.5121	.4879					.4879
3	.3696	.4947	.1357				.7660
4	.3111	.4748	.1926	.0215			.9245
5	.2830	.4600	.2181	.0370	.0019		1.0148
6	.2684	.4509	.2305	.0463	.0037	.0001	1.0663
Limit							1.0881

Unfortunately, the data from the current experiment are not well suited to exploring models of this type. As is apparent in Figure 16 there is a large practice effect. Since the model includes no provisions for practice effects, its expectation (and most peoples' intuition) would be that error rate would increase steadily until it reached a limiting value and would remain at this limit for the rest of the experiment. It is very difficult to apply this model to situations where practice effects are intermingled with inhibitory effects.

In summary, the model will certainly have to be extended if it is to provide a reasonable account of recognition memory when the foil is a new word. However, it is likely that a satisfactory model

can be formulated by appealing to a list-differentiation mechanism and to confusions about which alternative is represented by which event code. The development of a reasonable quantitative description will have to wait upon an experiment that can provide data on the development of inhibitory effects without being contaminated by concomitant practice effects.

If the basic ideas underlying the model are accepted, then it would be expected that the model should be easily extended to recall experiments. A natural assumption is that in recall the S selects a recent event code, and if more than one is recent he guesses among the set of recent codes. If this assumption is coupled with the assumption of a rapidly deteriorating object code and a list-differentiation mechanism, the model provides reasonable predictions for recall experiments. Using the parameter values estimated from the current experiment, the model predicts a rapid build-up of PI with an asymptotic probability of correct recall of .357 reached after 5 trials. Retention functions are reasonable in appearance, going from very high and nearly flat on the first trial to an asymptotic form which declines more precipitously to a lower level. Further, the model generates reasonable predictions about the relative frequency of intrusions of different source lag (cf., Figure 7). With the parameters estimated for the current experiment, about 70% of the intrusions would be from the immediately preceding item, 15% from the item 2 back in the sequence, 8% from the TBRI 3 back, and the remainder from all previous items. (No attempt was made to include extra-experimental intrusions in the analysis.)

Summary

The data of the current experiment strongly suggest that provisions for trace confusability will be an essential part of any account of recognition memory of the type demonstrated here. This conclusion in turn suggests that a similar assumption will have to be introduced into models for the recall situation. This is an important conclusion because many of the most popular memory models make no provision for difficulties in locating the trace of the TBRI. If these models are to be extended to situations where inter-trial PI is an important determinant of performance, it will be necessary to make some basic revisions in the postulated forgetting mechanisms.

Since there are very few quantitative models that assume trace confusability, a necessary first step in developing models is to isolate what appear to be the essential properties of a system which incorporates difficulties in locating the correct trace. It was proposed that two separate types of internal representations be postulated. On the one hand are "object codes," which contain information about the public meaning of an event. On the other hand are "event codes," which specify not only the characteristics of the event *per se*, but also information about when and where it occurred. It was suggested that trace confusability be identified as confusions among event codes. This scheme was then specialized to produce a model. The model assumes that event codes can be partitioned into only two classes, "recent" and "old." This assumption is combined with a very simple decision rule: Select

the alternative whose event code appears most recent, or, if they are both of the same age (both "recent," or both "old"), guess randomly.

The model was fitted to the data of the current experiment. It provided a credible description of performance on tests using intra-experimental foils, but it was not satisfactory when the foil was a word new to the experiment. There were two problems with new foils. First, the shape of the predicted retention function was in disagreement with the obtained function. Second, performance on tests using new foils was strongly dependent upon the number of prior trials, an effect for which the model makes no provision. Several alternative ways of modifying the model to make it more reasonable were suggested. While the experiment did not provide enough data to test these notions, they appeared to be reasonable. It is not overly optimistic to think that a model of the general form proposed here can provide a reasonable account of the data.

APPENDIX A

TABLED VALUES FOR SELECTED FIGURES

TABLE 1

OBSERVED AND EXPECTED RETENTION FUNCTIONS, TABLED VALUES FOR FIGURE 1

(DATA FROM PETERSON & PETERSON, 1959)

Retention Interval (Sec.)	Observed Value	Predicted Value
0	--	1.00
3	.77	.71
6	.56	.51
9	.31	.36
12	.22	.26
15	.12	.18
18	.08	.13

TABLE 2

OBSERVED AND PREDICTED RETENTION FUNCTIONS, TABLED VALUES FOR FIGURE 3

(DATA FROM HELLYER, 1962)

Number of Presentations		Retention Interval			
		3	9	18	27
8	O	.99	.89	.74	.66
	P	.98	.88	.76	.68
4	O	.94	.73	.56	.46
	P	.94	.73	.55	.47
2	O	.92	.54	.31	.22
	P	.89	.55	.30	.22
1	O	.89	.38	.21	.14
	P	.84	.40	.18	.13

TABLE 3

PREDICTED AND OBSERVED PI BUILD-UP FUNCTIONS OF FIGURE 6

(DATA FROM NOYD, 1965)

Trial		Retention Interval (Sec.) 4	8	24
1	O	.82	.67	.78
	P	.83	.73	.73
2	O	.48	.29	.25
	P	.49	.30	.25
3	O	.45	.30	.22
	P	.46	.26	.20
4-5-6	O	.46	.22	.15
	P	.43	.21	.15

TABLE 4

PREDICTED AND OBSERVED FREQUENCY OF INTRUSIONS

AS A FUNCTION OF SOURCE LAGE, FIGURE 10

(DATA FROM NOYD, 1965)

	Source Lag 1	2	3	4	5	6	7	8
Observed	.108	.019	.006	.005	.007	.007	.003	.002
Predicted	.108	.021	.007	.006	.005	.004	.003	.003

TABLE 5

ERROR RATE AS A FUNCTION OF FOIL LAG AND RETENTION INTERVAL, DATA OF FIGURES 11 AND 14

Foil Lag	8	20	30	Total
1	.207	.272	.302	.260
2	.117	.188	.228	.178
4	.102	.105	.176	.128
8	.077	.133	.142	.117
12	.046	.130	.157	.111
New	.046	.083	.080	.070

TABLE 6

POINTS FOR ROCs OF FIGURE 12 AND A_G OF FIGURE 13

Foil Lag		1	2	3	4	5	6	7	8	A_G
1	Hit	.490	.617	.688	.739	.790	.841	.905	1.000	
	F.A.	.095	.159	.210	.261	.312	.383	.510	1.000	.803
2	Hit	.573	.692	.766	.822	.859	.905	.307	1.000	
	F.A.	.051	.094	.140	.177	.233	.307	.426	1.000	.885
4	Hit	.623	.752	.825	.872	.913	.946	.973	1.000	
	F.A.	.027	.054	.087	.128	.175	.248	.377	1.000	.936
8	Hit	.618	.753	.815	.876	.913	.947	.977	1.000	
	F.A.	.024	.054	.088	.125	.186	.248	.383	1.000	.948
12	Hit	.620	.741	.832	.891	.925	.955	.982	1.000	
	F.A.	.019	.046	.076	.110	.169	.260	.381	1.000	.954
New	Hit	.666	.782	.855	.929	.961	.976	.986	1.000	
	F.A.	.013	.023	.038	.070	.144	.217	.333	1.000	.982

TABLE 7

ERROR RATE ON TESTS USING FOILS OF LAG 1 AND 2 AS A FUNCTION OF THE ACTUAL INTER-PRESENTATION INTERVAL, TABLED VALUES FOR THE DATA OF FIGURE 15, COMBINED AS INDICATED

Inter-Presentation Interval	Foil Lag 1	Foil Lag 2	
21	.269		
33	.253		
42		.259	
43	.238		
54		.139	
64		.190	.164
76		.222	
86		.134	.178
96		.111	

TABLE 8

PERFORMANCE ON TESTS USING NEW WORDS AS FOILS AS A FUNCTION OF TRIALS, TABLED VALUES FOR THE DATA OF FIGURE 16

1	2	3	4	5	6-25	16-25	26-35	36-45	46-61
.998	.862	.835	.835	.822	.822	.889	.945	.914	.945

APPENDIX B

THE SEQUENCES OF EXPERIMENTAL CONDITIONS

There are 6 main orders of TBRIs. Each one of these orders was assigned a single order of experimental conditions. Each stimulus list has associated with it conditions which represented one replication of a complete Foil Lag (6) by TBRI within-item serial position condition (3) by Foil within-item serial position condition (3). It is convenient to combine the last two conditions into a single within-item serial position condition with 9 levels. If (i,j) represents the $i^{\underline{th}}$ TBRI within-item serial position and the $j^{\underline{th}}$ Foil within-item serial position, then the mapping from the two within-item serial position conditions onto the combined condition, (i,j):k, is: (1,1):1, (1,2):2, (1,3):3, (2,1):4, (2,2):5, (2,3):6, (3,1):7, (3,2):8, (3,3):9. Table 1 shows the rules used to assign a retention interval dummy index to each combination of the foil lag and within-item serial position conditions, for each separate stimulus list. This scheme insures that every foil lag and within-item serial position condition will occur equally often with every retention interval index value. The index values were then associated with particular retention intervals so that every assignment of retention interval to dummy index would use each retention interval just once and so that every index value would occur equally often with every retention interval. The procedure used is displayed in Table 2.

TABLE 1

ASSIGNMENT OF RETENTION INTERVAL INDEX TO COMBINATIONS OF THE FOIL LAG AND WITHIN-ITEM SERIAL POSITION CONDITIONS FOR EACH STIMULUS LIST

List 1

Foil Lag	Within-Item Condition 1	2	3	4	5	6	7	8	9
1	c	b	a	a	c	b	b	c	a
2	a	b	c	b	a	c	c	a	b
3	a	c	b	b	c	a	c	b	a
4	b	a	c	c	a	b	a	b	c
5	b	c	a	c	b	a	a	c	b
6	c	a	b	a	b	c	b	a	c

List 2

Foil Lag	Within-Item Condition 1	2	3	4	5	6	7	8	9
1	c	a	b	a	b	c	b	a	c
2	c	b	a	a	c	b	b	c	a
3	a	b	c	b	a	c	c	a	b
4	a	c	b	b	c	a	c	b	a
5	b	a	c	c	a	b	a	b	c
6	b	c	a	c	b	a	a	c	b

List 3

Foil Lag	Within-Item Condition 1	2	3	4	5	6	7	8	9
1	b	c	a	c	b	a	a	c	b
2	c	a	b	a	b	c	b	a	c
3	c	b	a	a	c	b	b	c	a
4	a	b	c	b	a	c	c	a	b
5	a	c	b	b	c	a	c	b	a
6	b	a	c	c	a	b	a	b	c

List 4

Foil Lag	Within-Item Condition 1	2	3	4	5	6	7	8	9
1	b	a	c	c	a	b	a	b	c
2	b	c	a	c	b	a	a	c	b
3	c	a	b	a	b	c	b	a	c
4	c	b	a	a	c	b	b	c	a
5	a	b	c	b	a	c	c	a	b
6	a	c	b	b	c	a	c	b	a

List 5

Foil Lag	Within-Item Condition 1	2	3	4	5	6	7	8	9
1	a	c	b	b	c	a	c	b	a
2	b	a	c	c	a	b	a	b	c
3	b	c	a	c	b	a	a	c	b
4	c	a	b	a	b	c	b	a	c
5	c	b	a	a	c	b	b	c	a
6	a	b	c	b	a	c	c	a	b

List 6

Foil Lag	Within-Item Condition 1	2	3	4	5	6	7	8	9
1	a	b	c	b	a	c	c	a	b
2	a	c	b	b	c	a	c	b	a
3	b	a	c	c	a	b	a	b	c
4	b	c	a	c	b	a	a	c	b
5	c	a	b	a	b	c	b	a	c
6	c	b	a	a	c	b	b	c	a

TABLE 2

ASSIGNMENT OF RETENTION INTERVALS TO INDEX VALUES,

THE SIX CYCLES SHOWN WERE REPEATED THREE TIMES

Retention Interval Index	Six Cycles of Index Assignments					
	1	2	3	4	5	6
a	8	20	30	20	8	30
b	20	30	8	8	30	20
c	30	8	20	30	20	8

In Table 3 is one of the actual stimulus lists used (List 1) with associated Foil Lag and combined within-item serial position conditions. The first column of the Table contains the TBRI for that trial. The within-TBRI ordering of the words is the same here as it was during the presentation of the item to the subject. In the second column are a pair of numbers which represent, respectively, the Foil Lag and within-item serial position conditions. The numbering of the Foil Lag conditions is: (Lag 1):1, (Lag 2):2, (Lag 4):3, (Lag 8):4, Lag 12):5, (New Foils):6. The next column contains the retention interval index value for that trial. In the final column are the recognition alternatives, the target on the left and the foil on the right.

In Table 4 is the ordering of Foil Lag and within-item serial position conditions for all six stimulus lists. Across the separate lists the ordering of the TBRIs was in accordance with a balanced Latin square (see the discussion of the experimental design in Chapter II for more detail).

TABLE 3

LIST 1

1.	Gulf	Mind	Film	11	a	Gulf	Farm
2.	Name	Ache	Loaf	29	b	Loaf	Bead
3.	Shop	Dirt	Gang	25	a	Dirt	Ache
4.	Bark	Post	Knee	43	c	Bark	Shed
5.	Harm	Chin	Sash	42	c	Harm	Mind
6.	Wool	Task	Host	62	a	Task	Cord
7.	Lark	Maid	Weed	22	b	Lark	Chin
8.	Pile	Seed	Duke	48	b	Duke	Limb
9.	Camp	Joke	Hand	15	c	Joke	Seed
10.	Moon	Colt	Town	39	a	Town	Host
11.	Tube	Plan	Frog	44	c	Plan	Shop
12.	Beam	View	Note	13	a	Beam	Frog
13.	Leaf	Path	Mast	61	c	Leaf	Kite
14.	Cost	Bait	Stem	31	a	Cost	Moon
15.	Grip	Twin	Dome	59	b	Dome	Gang
16.	Mode	Hail	Coin	53	a	Mode	Knee
17.	Sage	Noon	Lung	63	b	Lung	Road
18.	Rail	Wolf	Food	51	b	Rail	Wool
19.	Chip	Bank	Rear	12	b	Chip	Wolf
20.	Tire	Crew	Life	16	b	Crew	Rear
21.	Arch	Pump	Wire	57	a	Wire	Camp
22.	Brow	Coat	Soap	52	c	Brow	Colt
23.	Desk	Heap	Foam	41	b	Desk	Grip
24.	Soil	Fork	Pace	56	a	Fork	Note
25.	Heal	Salt	Tile	17	b	Tile	Soil
26.	Rust	Gale	Hoof	36	a	Gate	Soap
27.	Lash	Edge	Port	38	b	Port	Heap
28.	Care	Wall	Meal	63	b	Meal	Song
29.	Bath	Mind	Sack	45	a	Mind	Pump
30.	Plum	Dock	Coat	62	a	Dock	Glue
31.	Mill	Crab	Soup	46	b	Crab	Foam
32.	Comb	Show	Lack	21	a	Comb	Plum
33.	Rose	Mess	Cane	62	a	Mess	Cart
34.	Harp	Dust	Gown	33	b	Harp	Coat
35.	Glow	Bean	Hawk	23	c	Glow	Cane
36.	Beak	Well	Nail	28	c	Nail	Dust
37.	Palm	Gold	Wing	34	b	Gold	Rose
38.	Lock	Part	Dawn	54	c	Part	Rust
39.	Doll	Nook	Lamp	24	b	Nook	Palm
40.	Shoe	Fire	Tend	49	c	Tent	Lace
41.	Bird	Trip	Rice	42	a	Bird	Mess
42.	Rope	Wage	Barn	19	a	Barn	Rice
43.	Bent	Jail	Doom	37	c	Doom	Doll
44.	Spot	Milk	Bush	61	c	Spot	Band
45.	Yarn	Pool	Heal	27	c	Heat	Bent
46.	Room	Dove	Pier	47	a	Pier	Lock
47.	Fish	Lime	Goal	35	c	Lime	Jail
48.	Moss	Boat	Tree	55	b	Boat	Well
49.	Fate	Belt	Mark	14	a	Fate	Tree
50.	Rope	Pipe	Lawn	63	b	Lawn	Wind
51.	Sand	Grin	Cape	26	c	Grin	Mark
52.	Boot	Flag	Gift	61	c	Boot	Seal
53.	Mule	Twig	News	58	c	News	Trip
54.	Case	Horn	Tail	18	c	Tail	Twig

TABLE 4

SEQUENCES OF EXPERIMENTAL CONDITIONS

	List					
Trial	1	2	3	4	5	6
1	11 c	19 c	15 b	63 b	26 b	32 a
2	29 b	21 c	19 b	29 b	39 b	34 c
3	25 a	29 a	41 a	42 b	46 c	22 c
4	43 c	63 a	31 c	62 c	41 c	16 c
5	32 c	62 c	49 b	25 b	29 c	11 a
6	62 a	35 a	35 c	19 c	58 c	46 a
7	22 b	58 b	59 a	49 a	12 c	52 a
8	48 b	41 a	17 a	53 c	19 a	62 b
9	15 c	51 b	55 c	55 a	45 b	38 b
10	39 a	15 b	47 c	14 c	63 c	61 c
11	44 c	48 b	57 c	58 a	61 a	59 c
12	13 a	52 a	34 a	21 b	55 c	43 a
13	61 c	12 a	21 c	12 a	11 a	12 b
14	31 a	23 a	62 a	41 c	31 b	49 b
15	59 b	61 b	63 c	32 a	25 a	21 a
16	53 a	62 c	29 c	33 b	52 b	25 c
17	63 b	36 c	61 b	39 c	36 a	62 b
18	51 b	43 b	25 b	26 a	63 c	35 a
19	12 b	18 a	23 b	61 a	15 c	53 b
20	16 b	31 a	43 c	27 a	62 b	63 a
21	57 a	42 c	39 a	45 c	17 c	55 b
22	52 c	16 c	11 b	36 c	42 a	19 b
23	41 b	46 a	61 b	24 c	43 b	54 a
24	56 a	24 a	63 c	62 c	61 a	28 b
25	17 b	32 b	52 c	11 b	35 b	15 a
26	36 a	56 b	44 b	52 b	63 c	24 b
27	38 b	11 c	32 b	31 c	21 b	23 b

TABLE 4 (Continued)

SEQUENCES OF EXPERIMENTAL CONDITIONS

Trial	List 1	2	3	4	5	6
28	63 b	59 c	12 c	15 a	38 c	63 a
29	45 a	63 a	63 c	35 b	51 c	41 b
30	62 a	49 a	28 a	16 b	54 a	33 c
31	46 b	63 a	24 a	59 b	49 c	13 c
32	21 a	34 b	33 a	54 b	18 b	37 a
33	62 a	25 c	58 b	46 b	57 b	47 a
34	33 b	37 c	37 b	63 b	24 c	44 c
35	23 c	54 c	14 c	62 c	27 a	61 c
36	28 c	22 b	22 a	47 b	32 c	45 b
37	34 b	39 b	45 a	37 b	16 a	48 c
38	54 c	45 c	26 c	17 a	44 a	62 b
39	24 b	53 c	56 a	28 c	14 b	57 b
40	49 c	13 b	62 a	43 a	62 b	29 a
41	42 a	44 b	48 a	57 c	53 a	56 c
42	19 a	47 c	62 a	44 a	47 b	27 c
43	37 c	57 a	13 a	13 c	34 c	42 c
44	61 c	14 a	38 c	18 b	48 a	58 a
45	27 c	28 c	46 c	34 a	62 b	31 b
46	47 a	17 b	61 b	22 c	56 b	51 c
47	35 c	33 c	36 b	23 a	13 b	39 c
48	55 b	55 a	16 a	51 a	37 a	26 a
49	14 a	62 c	54 b	63 b	23 c	17 c
50	63 b	38 a	53 b	61 a	28 b	61 c
51	26 c	61 b	27 b	38 a	61 a	63 a
52	61 c	27 b	42 b	56 c	59 a	18 a
53	58 c	26 b	18 c	48 c	33 a	14 b
54	18 c	61 b	51 a	61 a	22 a	36 b

REFERENCES

Atkinson, R. C., and Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), _The psychology of learning and motivation_, Vol. 2. New York: Academic Press, 1968.

Bower, G. A multicomponent theory of the memory trace. In K. W. Spence and J. T. Spence (Eds.), _The psychology of learning and motivation_, Vol. 1. New York: Academic Press, 1967.

Brown, J. Information, redundancy and decay of the memory trace. In _The mechanization of thought processes_. London: Her Majesty's Stationary Office, 1959.

Conrad, R. Serial order intrusions in immediate memory. _British Journal of Psychology_, 1960, _51_, 45-48.

Conrad, R. Interference or decay over short retention intervals? _Journal of Verbal Learning and Verbal Behavior_, 1967, _6_, 49-54.

Green, D. M., and Swets, J. A. _Signal detection theory and psychophysics_. New York: Wiley, 1966.

Hellyer, S. Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. _Journal of Experimental Psychology_, 1962, _64_, 650.

Johnson, N. F. The role of chunking and organization in the process of recall. In Bower, G. (Ed.), _The psychology of learning and motivation_, Vol. 4. New York: Academic Press, 1970.

Keppel, G., and Underwood, B. J. Proactive inhibition in short-term retention of single items. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 153-161.

Krantz, D. Threshold theories of signal detection. Psychological Review, 1969, 76, 308-324.

Loess, H. Proactive inhibition in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 362-368.

Loess, H. Short-term memory, word class, and sequence of items. Journal of Experimental Psychology, 1967, 74, 556-561.

Loess, H. Short-term memory and item similarity. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 87-91.

Loess, H., and Waugh, N. Short-term memory and intertrial interval. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 455-460.

Luce, R. D. Individual choice behavior. New York: Wiley, 1959.

Luce, R. D., and Galanter, E. Discrimination. In R. D. Luce, R. R. Bush, and E. Galanter, (Eds.), Handbook of mathematical psychology, Vol. 1. New York: Wiley, 1963.

Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. Psychological Review, 1968, 75, 421-441.

Martin, E. Verbal learning theory and independent retrieval phenomena. Psychological Review, 1971, 78, 314-332.

Melton, A. W. Sameroff, A., and Schubot, E. D. Short-term recognition memory. Human Performance Center Memorandum Report No. 2, The University of Michigan, Ann Arbor, 1967.

Norman, D. A., and Rumelhart, D. E. A system for perception and memory. In D. A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970.

Noyd, D. Proactive and intrastimulus interference in short-term memory for two-, three-, and five-word stimuli. Paper presented at the Annual Meeting of the Western Psychological Association, Honolulu, Hawaii, 1965.

Peterson, L. R., and Peterson, M. J. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), Verbal learning and verbal behavior. New York:

Shepard, R. N., and Chang, J. Forced-choice tests of recognition memory under steady-state conditions. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 93-101.

Shepard, R. N., and Teghtsoonian, M. Retention of information under conditions approaching a steady state. Journal of Experimental Psychology, 1961, 62, 202-222.

Spence, K. W. The nature of discrimination learning in animals. Psychological Review, 1936, 43, 427-449.

Torgerson, W. S. Theory and method of scaling. New York: Wiley, 1958.

TECHNICAL REPORTS NOT OTHERWISE PUBLISHED

1. Phillips, L. D. Some components of probabilistic inference. January 1966.

3. Biederman, I. Human performance in contingent information processing tasks. October 1966.

8. Ligon, E. The effects of similarity on very-short-term memory under conditions of maximal information processing demands. May 1968.

9. Triggs, T. J. Capacity sharing and speeded reactions to successive signals. August 1968.

11. Lively, B. L. The Von Restorff effect in very-short-term memory. December 1968.

13. Swensson, R. G. The elusive tradeoff: Speed versus accuracy in choice reaction tasks with continuous cost for time. December 1968.

15. Kamlet, A. S. Processing of sequentially presented signals in information combining tasks. June 1969.

16. Pollatsek, A. W. Rehearsal, interference, and spacing of practice in short-term memory. July 1969.

17. Garskof, M. H. The effect of spacing and variation of repetition in short-term memory. August 1969.

19. Du Charme, W. M. A responsive bias explanation of conservative human inference. December 1969.

20. McCormack, P. D. Monitoring eye movements during the learning of paired-associate lists. March 1970.

21. Goodman, B. C. Risky decisions by individuals and groups. June 1970.

22. Wattenbarger, B. L. The representation of the stimulus in character classification. August 1970.

23. Gardner, G. T. Spatial processing characteristics in the perception of brief visual arrays. August 1970.

24. Adams, R. A. S. Interference in short-term retention of discrete movements. September 1970.

25. Armstrong, T. R. Feedback and perceptual-motor skill learning: A review of information feedback and manual guidance training techniques. August 1970.

26. Armstrong, T. R. Training for the production of memorized movement patterns. August 1970.

27. Collins, A. M. The effect of visual stimulus traces on memory. August 1970.

28. Cohen, V. V. R. Short-term memory for quantitative information from three kinds of visual displays. June 1971.

29. Gelfand, H. Organization in free recall learning: Output contiguity and interresponse times as a function of presentation structure. June 1971.

30. Roberts, K. H. An investigation of paraphrasing: The effects of memory and complexity. June 1971.

MEMORANDUM REPORTS AFTER 1968

7. Edwards, W. A bibliography of research on behavioral decision processes to 1968. January 1969.

9. Greeno, J. G. A cognitive interpretation of negative transfer and forgetting of paired associates. November 1969.

10. Martin, E. Associative interference theory and spontaneous recovery. November 1969.

12. Greeno, J. G. Theoretical entities in statistical explanation. October 1970.

13. Greeno, J. G. Technical and informal theories about mental entities. October 1970.